ARQ Systems for Wireless Communications

Dem Fachbereich
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
vorgelegte und genehmigte

Dissertation

von
Michael Philipp Schmitt M.S. SUNY/Buffalo
geboren am 28. Januar 1972 in Aschaffenburg

Referent:	Prof. Dr.-Ing. B. Dorsch
Korreferent:	Prof. Dr.-Ing. R. Jakoby
Tag der Einreichung:	15. Juni 2002
Tag der mündlichen Prüfung:	26. September 2002

D17
Darmstädter Dissertation
Darmstadt 2002

To my mother

Agnes Schmitt

and
to the memory of my father

Alfred Schmitt

Acknowledgements

Writting a dissertation takes its time. Many hours are involved - including those which had to be borrowed from my private life. Therefore, I want to take the opportunity to thank my family, my friends, and, of course, Heidi for their patience and their support throughout the years.

But time is not all it takes. The ideas which led to this work grew from many interactions. Therefore, I would like to thank my former colleagues at Panasonic European Laboratories, Langen Germany, especially Christian Wengerter and Thomas Wiebke. Many ideas, which influenced this work considerably, originated from our collaboration. Also, I would like to thank Dr. Jagath-Kumara who made my stay at the Department of Electrical Engineering at Massey University, Palmerston North, New Zealand possible. Large parts of this work were written in the comfortable atmosphere, he provided. And, of course, thanks to all my colleagues at the Institut für Nachrichtentechnik at Technische Universität Darmstadt. The years with you will not be forgotten.

Finally and most of all, I would like to thank Prof. Dr. Bernhard Dorsch for supporting and encouraging me. I deeply value the freedom he granted to allow myself to concentrate on my favorite area of interest. I do not take this for granted.

Dachau, June 2002

Michael P. Schmitt

Kurzfassung

Drahtlose digitale Kommunikation leidet immer unter kanalbedingten Übertragungsfehlern. In den vergangenen Jahrzehnten wurden verschiedene Typen von Vorwärtsfehlerkorrekturverfahren (FEC) entwickelt, um die Fehlerhäufigkeit zu mindern. Den hohen Anforderungen an die Bitfehlerrate bei der Übermittlung von Daten kann jedoch mit FEC - speziell bei stark variierender Kanalgüte - nicht Rechnung getragen werden. Dafür eignen sich besonders Fehlerkorrrekturverfahren, bei dem der Empfänger den Sender über den fehlerfreien oder -behafteten Empfang der Daten informiert und dieser gegebenenfalls einen neuen Übertragungsversuch startet. Diese sogenannten *Automatic-Repeat-Request* (ARQ) Verfahren sind der Hauptbestandteil der vorliegenden Arbeit. Die neuen Beiträge zu diesem Gebiet lassen sich in drei Teile gliedern.

Der erste Beitrag liegt auf dem Gebiet der ARQ Klassifikation. Die Klassifikation, wie sie sich in der vorhandenen Literatur wiederfindet, ist inkonsistent und basiert zum Teil auf speziellen Decoderrealisierungen. Dies ist jedoch ein falsches Konzept einer Verfahrensklassifikation, da sie nur auf Eigenschaften des Verfahrens basieren soll und nicht darauf wie ein spezieller Decoder sie verwertet. Aus diesem Grund wird in dieser Arbeit für ARQ Verfahren eine eindeutige Definition gegeben und sie von anderen Wiederholungsverfahren unterschieden, bevor eine auf Verfahrenseigenschaften basierende Klassifikation eingeführt wird.

Der zweite Beitrag dieser Arbeit liegt auf dem Gebiet der ARQ Systemeigenschaften. Von der Vielzahl möglicher Eigenschaften wird hauptsächlich der Durchsatz (*Throughput*) in der Literatur behandelt, für den es zudem eine Vielzahl unterschiedlicher Definitionen gibt. Aus diesem Grund wird eine einheitliche Definition des Durchsatzes als Mass für die informationstheoretische Leistungsfähigkeit eines ARQ Verfahrens gegeben. Als eine Konsequenz dieser Definition dient die bekannt Kanalkapazität als obere Grenze. Darüber hinaus wird der Durchsatz kaum analytisch behandelt, mit Ausnahme der ineffizienten gedächtnislosen Systeme. In der Arbeit wird, ausgehend von sogenannten Ablehnungs-wahrscheinlich-keiten, der Durchsatz ermittelt. Neben dem Durchsatz werden andere wichtige Systemeigenschaften, wie die maximale und durchschnittliche Anzahl der Übertragungen, verschiedene Verzögerungen und die effektive Datenrate eingeführt. Diese Eigenschaften werden sukzessive von bereits bestehenden Systemeigenschaften und weiteren Systemparametern hergeleitet und mittels der erwähnten Ablehnungswahrscheinlichkeiten beschränkt.

Als letzter Beitrag dieser Arbeit wird das optimale Kombinieren von ARQ Übertragungen zur Vorbereitung auf das ML oder MAP Decodieren hergeleitet. Hierbei wird von einer variierenden Abbildungsfunktion der Codewörter auf die Übertragungssymbole ausgegangen.

Abstract

Wireless digital communication always suffers from errors introducted by the channel. Throughout the years, various types of forward error correcting schemes have been developed to soften this effect. However, the recently increasing demands of transmitting data very reliably, i.e. with almost neglectable bit error rate, also over time varying channels, called for other error control systems, which inform the transmitter of the correct or incorrect reception of data. These systems are called automatic repeat request schemes (ARQ) and are the main topic of this work. The contributions of this work to this area are threefold.

Firstly, classifications of ARQ systems which can be found in literature are inconsistent and, even worse, the most popular ones are based on actual decoder realizations. This, however, represents a wrong concept, since a system classification should be based on system properties, such as what is sent for the n-th transmission, and not on how a specific realization makes use of it. Therefore, this work, first of all, distinguishes ARQ system from other feedback systems and proceeds with a classification scheme which is purely based on ARQ system properties. This classification is used throughout the remainder of this work.

The second contribution lies in the area of ARQ performance measurements. From the realm of these measures, mainly one is treated in literature, namely the throughput. In addition to the fact that there are almost as much throughput definitions as there are ARQ papers, it is hardly treated analytically. The only exceptions are the inefficient memoryless ARQ schemes, which, due to their lack of memory, are much simpler to analyze. Therefore, this work introduces a definition for the throughput as a measurement for the information theoretical performance of ARQ systems, which is independent of the transmission protocol (Stop-and-Wait, Go-Back-N, Selective-Repeat). As a consequence, the channel capacity serves as ultimate performance bound of this throughput. Besides the throughput, other important ARQ system properties such as maximum and average number of transmissions, various delays, and the data rate are introduced. All these measures are derived successively from another, including more and more system and environment parameters. All measurements are bounded with the help of the set of rejection probabilities. Using the results of this work and the set of rejection probabilities of a certain ARQ system and channel model, a detailed insight into the performance of the system can be obtained.

Finally, the problem of how to combine independent transmissions of a codeword with possibly varying mapper is investigated. More specifically, optimal combining for maximum likelihood, as well as MAP decoding is investigated for a variety of wireless channel models.

Erklärung

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation alleine und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Dachau, den 23.05.2002

Michael. P. Schmitt

Lebenslauf

1988			Realschulablschluss
1988	-	1991	Berufsausbildung zum Industrieelektroniker, YMOS AG, Obertshausen
1991			Gesellenprüfung Industrieelektroniker
1991	-	1992	Staatliche Fachoberschule Aschaffenburg
1992			Allgemeine Fachhochschulreife
1992	-	1993	Grundstudium der Elektrotechnik, FH Darmstadt
1993			Fachgebundene Hochschulreife
1993	-	1995	Grundstudium der Elektrotechnik, TU Darmstadt
1995	-	1996	Hauptstudium der Theoretischen ET, TU Darmstadt
1996	-	1998	Masterstudium der Digitalen Kommunikation, SUNY Buffalo, NY, USA
1998			Master of Science in Electrical Engineering
1998	-	2001	Promotionsstudium im Rahmen des GK ISIA, TU Darmstadt
2001	-		Systemingenieur bei EADS TELECOM, Unterschleissheim

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As the title suggests, this work is concerned with a special class of error control mechanism - ARQ systems - for the use over wireless radio links. The introduction chapter aims to precise the scope of this work and show its relation to the ISO-OSI model of communication systems. Also, a reasoning for the type of investigated systems is provided and the structure of the remainder of this work is given.

Therefore, Section 1.1 introduces the mentioned ISO-OSI model and highlights the relevant layers for this work in more detail. Sections 1.2 and 1.3 respectively introduce the two basic principles of error control, namely forward error correction (FEC) and automatic repeat request (ARQ). Finally, a convergence of the two lowest ISO-OSI layers are proposed.

1.1 The ISO-OSI Model

Digital communications systems have become highly complex systems with specifications filling shelves. Not only to reduce their design complexity, but in order to make their development at all possible, their underlying communication process is organized as a series of layers, each one built upon the other: Layer n on one communication terminal carries on a conversation with the corresponding layer on the other terminal. The rules and conventions used in this conversation are known as the layer n protocol.

The International Standards Organization (ISO) has proposed the so-called ISO-OSI (Opens Systems Interconnection) Reference Model as a generic model for network (communication) systems, specifying the individual layers and what they should do. However, the model does not specify the exact services and the protocols for each layer. Figure 1.1 depicts the ISO-OSI reference model (in the following called OSI model). Seven layers are defined from which only the first two layers - the physical and the data link layer - are of concern for this work and therefore only these will be discussed in more detail later. [Tan96] provides a good introduction to the different layers of the OSI model.

As Figure 1.1 suggests, the corresponding entities on the same layer of the communication

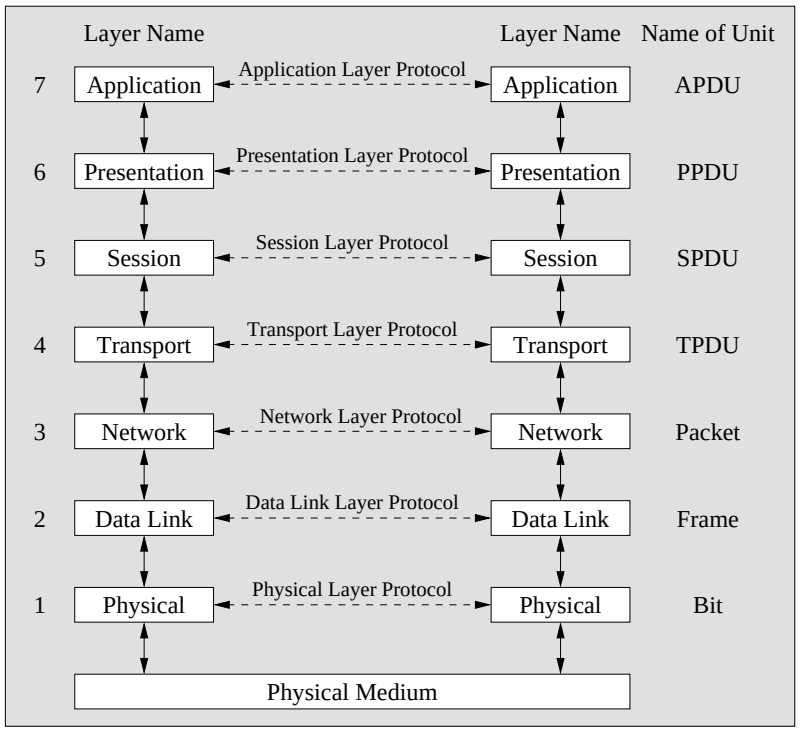

Figure 1.1: The ISO-OSI Reference Model

terminals (called peers) are communicating which each other according to their specified protocol via the dotted virtual communication channels. Obviously, in order to do so, they do not actually exchange data via these lines. Instead, at the transmitter side, each layer passes its information to the layer immediately below it, until the lowest layer is reached. The physical layer accesses the physical medium in which the actual communication takes place. At the receiver, the information is then passed from the lowest layer upward to the higher layers.

Between each pair of adjacent layers there is an interface, defining the services and the operations to access these services, which the lower layer offers to the upper one. Herein lies the advantage for the design process: As soon as an agreement on the services which each layer should provide and the coresponding service interface and protocol is reached, each layer can be developed independently of the other layers.

At a typical interface, the upper layer $n+1$ passes its data to be transmitted to the lower layer n. This data unit is called the SDU (Service Data Unit) of layer n. In order to transfer the SDU and depending on the size of the SDU, the layer n either has to split it into several pieces or merge several SDUs to a new entity. On each newly created entity, a header is appended and passed as a separate unit, called PDU (Protocol Data Unit), to next lower layer $n-1$ (becoming the SDU of layer $n-1$). The header in each layer n PDU is required by the layer n peer to carry out the protocol and reassemble/split the layer n SDU. The important principle of the layer model is that it appears to the layer $n+1$, as if it transmits its PDUs in one piece via the virtual layer $n+1$ peer

to peer connection.

Figure 1.1 also shows the specific names of the units, which are exchanged via the virtual connection in each layer. Specifically, on the network layer the PDUs are called packets, on the data link layer frames and on the physical layer they are usually known as bits. In Section 1.4, however, it is reasoned that the PDUs of the layer 1 are not necessarily bits.

The services, each layer provides, can be classified according to their connection type. In a connection-oriented service, a connection is established, PDUs are transmitted in sequence (PDU pipe), and finally the connection is released. In contrast, a connectionless service does not set up a specified connection. Instead all PDUs are routed individually and may therefore arrive out of sequence. In addition to the connection type, a service, which a layer offers to the one above, can be characterized by its quality. In literature reliable and unreliable services are distinguished. In this context, an unreliable service tries to transmit the data with a high probability of arrival, but does not guarantee its correct arrival or even its arrival at all (best effort service). On the other hand, a reliable services never loses data. This is achieved via receiver acknowledgments. In this case, an unacknowledged PDU is retransmitted, introducing additional overhead and delays.

In the following, the two layers of the OSI model and their services which are important for this work are discussed. Both layers deal solely with a point-to-point communication and can therefore provide a connection oriented service only.

The Physical Layer is concerned with the transmission of raw information over the physical medium which lies below this layer and everything which is involved to solve the arising realm of problems. In order to do so, it is responsible for the initial set-up of the physical connection, for solving the various synchronization problems (such as symbol timing, carrier synchronization, etc.), for the mapping of the layer 1 PDUs onto appropriate channel symbols (modulation) and their recovery (demodulation), and all related issues. Consequently, the terminology bit for the layer 1 PDU refers to a certain although predominant, realization. With respect to the ideas of the ISO-OSI layer model the more general terminologies transmission or channel symbols are used. Associated with this context, as it will be discussed in Section 1.4, are also the so-called transmission codes, i.e. codes which are an inherent part of the transmission process, such as softdecision decoded convolutional codes or coded modulation.

The methods, which are utilized to accomplish this task, heavily depend on the transmission media, i.e. the physical communication channel type with its characteristics. Typical channels are copper wire (twisted pair, coaxial cable), optical connections (fiber) and wireless channel (infra red, radio). The problems which arise and their proposed solutions are hardly discussed in computer network literature since it is the topic of a complete engineering science, namely digital communications with all its related aspects. From the mentioned transmission media, the wireless communication channel constitutes the biggest challenge due to its highly unreliable character. Unreliability in that context means that the transmission is occasionally erroneous: transmitted symbols are corrupted, frames are lost, or the connection is lost at all.

Hence, the design issue of the physical layer is to provide a high data rate at an acceptable error

rate and the only service this layer offers to the data link layer is a connection-oriented unreliable raw data transmission. This service is usually referred to as a bit pipe and the PDU of layer 1 as a bit. Especially in optical transmissions, multilevel symbols formed by grouping several bits are used for the transmission and the layer 1 PDUs will be called transmission symbols. Multilevel modulation, as this form of data modulation is known, is also popular for the transmission of digital data over the wireless channel to increase the data rate.

The Data Link Layer has two main tasks. Firstly, it controls the access to the transmission media. This media access control (MAC) becomes necessary if entities share the same physical channel. Well known MAC principles for the wireless channel are frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA). All of these MAC procedures require some form of synchronization between the entities while the ALOHA MAC procedure allows the entities to transmit data whenever they want. If the physical dimensions of the transmission media are small, such as in local area computer networks, carrier sense multiple access (CSMA) can be used.

The second main task of the data link layer arises from the unreliable physical layer. As mentioned, despite the arrangements made in layer 1 to keep the bit error rate (BER) as low as possible, residual errors will still occur to a certain amount. In this context, the data link layer is responsible for providing a service interface with a defined set of quality measurements to the network layer. These set of quality measurements can include the data rate (fixed or variable, minimum and maximum value,...), the residual BER (variable or maximal value), the introduced delay (fixed or variable, maximum value), the order of the SDUs (order preserved or reordering allowed), and so on. Despite the fact that the layer 2 must make use of the unreliable service of layer 1 without any QoS, it has to apply error control methods in order to provide a reliable service.

As a result, the PDUs received from the network layer must be segmented or merged into appropriate sizes so that they can be encoded, appended with a header containing layer 2 protocol information (especially information about the boundaries of the frames) and passed to the physical layer.

If an unreliable service is provided, error detection can be used to inform the network layer about the erroneous PDU (frame) and leave the error handling to this layer. Also, forward error correction (FEC) methods can be used to enhance the degree of reliability, i.e. to lower the error rate. Section 1.2 discussed these methods in more detail.

However, in order to provide a reliable service, FEC methods are not sufficient in some cases. Instead, the transmitter is provided with a decision feedback about the correctness of each individual sent frame. If it is received correctly, a positive acknowledgment (ACK) is sent back, and on the other hand, the reception of a negative acknowledgment (NAK) indicates that something went wrong with that particular layer 2 PDU. Error control schemes with this type of acknowledgment procedure are called Automatic Retransmission Request (ARQ) schemes. An introduction to this systems is given in Section 1.3. This subpart of the data link layer is known as the logical link control (LLC). With this terms we can state the main topic of this work:

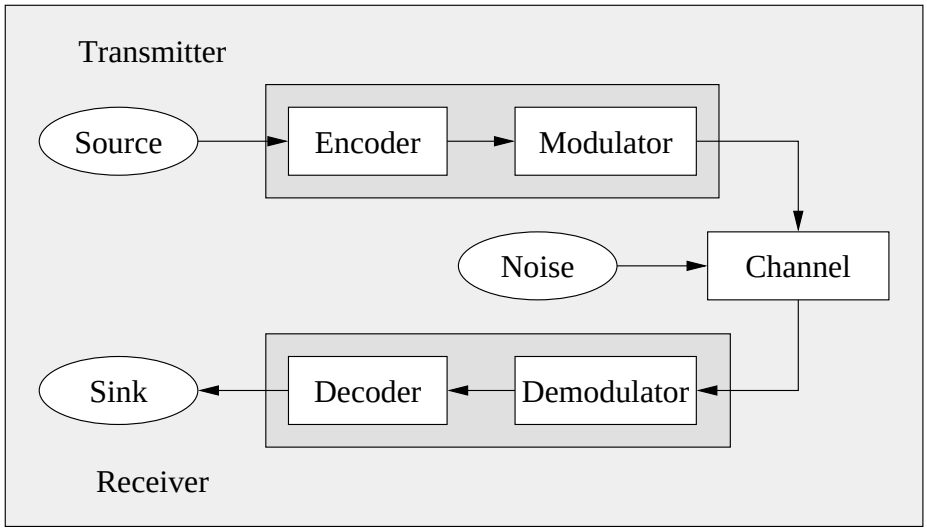

Figure 1.2: Block Diagram of a Forward Error Control Scheme.

Principles of efficiently providing a reliable service for the data link layer of wireless communication systems with the help of ARQ principles is the main topic of this thesis. Throughout this work an ideal MAC is assumed and all errors occurring are due to the impairment of the physical channel.

1.2 Forward Error Control Coding

Above it was mentioned, that purpose of the data link layer is to provide well defined services with certain quality to the network layer. To fullfill this task, despite the underlying unreliable physical layer, it makes use of error control methods, such as forward error control.

Forward error control can be subdivided into error detection and error correction schemes. However, the basic principle is common to both types. At the transmitter side, the data to be transmitted is encoded, i.e. known redundancy is added in form of additional bits. The encoded sequence is then modulated and transmitted over a noisy channel. At the receiver side, the received sequence is demodulated and passed to the decoder. The decoder uses the redundancy in the encoded sequence for either error detection (error detection codes) or even tries to correct the errors (error correction code). This method is schematically depicted in Figure 1.2.

In the following, the principle of how the redundant bits are used to detect and correct is outlined. Suppose we want to transmit k independent information bits. Hence, there are 2^k of those information words. Without a code, these bits are modulated, transmitted and demodulated on a bit basis. With the decision for the last bit, also a decision for one of the 2^k information words is made. With forward error control, the encoder adds $n - k$ additional bits so that the encoded word has a length of n bits. Although the length has increased, there are still only 2^k words, since the additional bits contain no additional information - they are derived from the

information word and are consequently redundant. The codewords are also transmitted over the channel and demodulated at the receiver. Of course, since the codewords have a length of n and the channel arbitrarily alters bits, the set of possible receive words has an increased size of 2^n compared to the uncoded transmission. This represents the key to error detection and correction. For error detection, the decoder simply checks if the received codeword is one of 2^k valid ones. If so, it assumes that no error has occurred, otherwise, an error is detected. The limitation of this approach has become also clear: If the channel alters a codeword into another valid one, the error is not detected. Hence, the design issue for the code is to make the set of codewords as distinct as possible, so that the event of an undetected error becomes rare. For error correction some kind of metric must be defined which measures how close an invalid received word is to any of the valid codewords. A common metric is the Hamming distance which is the number of differing bits. In case of an erroneously detected receiveword (received word), this metric is used to find the closest valid codedword and the decoder decides for that particular codeword as the one which was transmitted. Similar to the error detection, this procedure fails if the erroneously received word is a valid codeword (undetected errors), but it also fails when the receiveword is closer to another codeword than the transmitted one (decoding error). Again, the design issue is to make the codewords look as distinct as possible in the sense of that particular metric. In practice, the task of finding the closest codeword is a difficult venture, especially for codes with long codewords. As a consequence, theory of forward error control is mainly concerned with 2 areas:

1. Code design, i.e. the development of principles for creating a set of codewords with good distance properties.

2. Design of practical decoding algorithm, which allow the locating of the closest codeword in a tolerable time and with tolerable complexity.

History has shown that these two areas are usually treated by different communities and, moreover, sometimes during different decades. As an example, the Reed-Solomon Codes was first introduced in the ninteen-fifties, but it took as long as 1968 until an acceptable decoding algorithm was presented by Berlekamp.

It was mentioned in the beginning of this section, that the methods of FEC can be used to improve the QoS the data link layer offers to the network layer. The first thing, which should be mentioned in this context, is that this service improvement does not come for free. With FEC, only a specific fraction of the transmitted bits indeed carries information. This rate and its consequences for the data rate and the residual error rates are discussed in Section 4.1. Secondly, the FEC methods can in general be used in every layer. Nevertheless, incorporating FEC in layer 2 provides some advantages over FEC in higher layers. This issue will be treated in Section 1.4.

The error correction codes can be divided into two classes, depending on the way the redundancy is generated and added. The first class are the block codes, where a block of k information bits is used to algebraically compute a codeword of length n (hence the name (n, k) block code). This procedure is repeated with each new information block, whereby the computation is done

independently of the previous decoding operations. Typical information block lengths are larger than 100 $bits$. For the second class, the convolutional codes, the encoder has k parallel inputs and n parallel outputs. For the encoding procedure, the information sequence is multiplexed into k parallel streams which each one bit entering the encoder at a time step. From this k bits and a certain number v of previously inserted information bits the n output bits are computed and finally demultiplexed into the encoded bit stream. Hence, contrary to block codes, the encoding process for convolutional codes has a memory. The output sequence for each output can be mathematically described as the sum of k convolutions of the k input sequences with k so-called generator sequences, which explains the name for that class. Hence, the overall output is fully described by $k \cdot n$ generator sequences. k is typically 1 but also higher values up to 4 are in use. n typically ranges from $k + 1$ to 5.

The amount of literature dealing with error control coding seems unlimited, which was nicely illustrated by Wicker in his preface of [Wic95], "At this moment I have 25 books in my office that deal primarily with error control coding. I probably own more than that but my students have a habit of wandering off with them...".

Since the main emphasis on this work is on ARQ systems, the reader is referred for further details to any of this literature. The mentioned book provides a good introduction and in [Bos98] a more detailed treatment of error control codes is presented. The aspects of FEC, which are important for this work, will be illustrated in Chapter 5.

1.3 Automatic Repeat Request

As mentioned in the previous section, forward error correction has the disadvantage of possible decoding failures. The probability of this event can be made arbitrary small at the expense of an increasing amount of redundancy to be added. Especially, if the channel quality is highly varying the percentage of information contained in a codeword can become very small, since the amount of redundancy must be designed for the expected worst case. However, this results in an overdesigned system for better channel conditions.

For these cases a different solution has been proposed: Automatic Repeat Request (ARQ). In its pure form, the receiver makes use of an error detection code introduced by the transmitter in order to determine errors which possibly occurred during the transmission process. Then, just like in a regular forward error detection scheme, the codeword is transmitted over a noisy channel, demodulated, and decoded. In the case of a detected error, the erroneous codeword is discarded and a negative acknowledgment (NAK) is sent over a feedback channel to inform the transmitter of the erroneous transmission (see Figure 1.3 for a block diagram). Then, the transmitter takes action by retransmitting the codeword. This procedure continues until no error is detected and the transmitter is informed of that event via a positive acknowledgment (ACK). Thereafter the information is extracted by the decoder and passed to the sink.

Clearly, this principle has its limitations in the capability of the inherent error detection code. The residual error rate of the ARQ system is as high as the residual error rate of this code arising

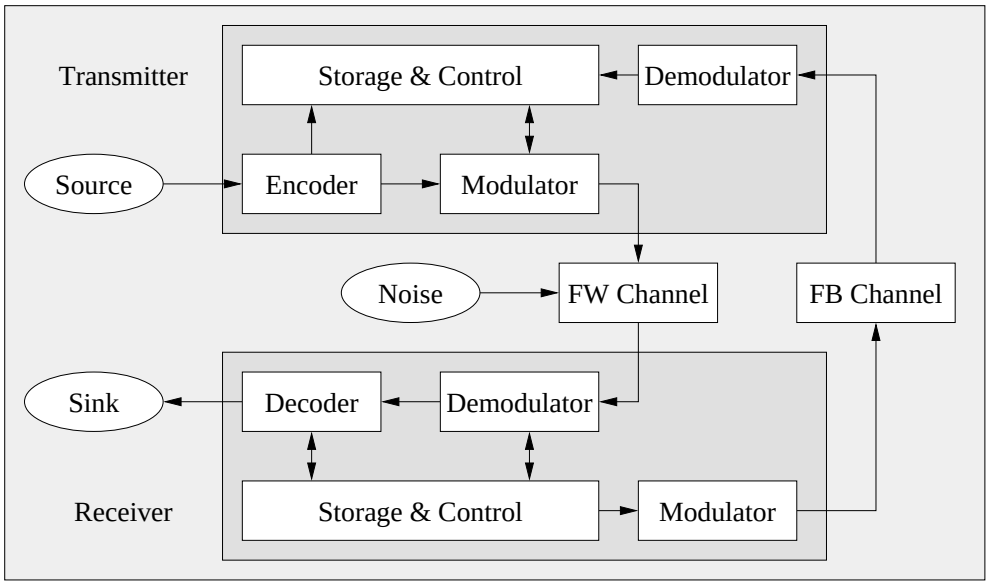

Figure 1.3: Block Diagram of an Automatic Repeat Request System.

from undetected errors. Also, from the ARQ principle it is clear that the system introduces a varying delay which might be unacceptable for some applications. And, of course, the principle can only be implemented if a feed-back channel is existent.

The ARQ principle described so far is known in literature as Pure ARQ and has an additional problem: Since it simply retransmits repetitions of the codeword and bases its decoding attempt purely on the current received version (the old ones were discarded), the system starts congesting if the error probability remains on a permanent high level. In that case, the effective data rate approaches zero and moreover the delay until the information reaches the sink increases infinitely.

As a solution to this problem so-called Hybrid ARQ (HARQ) systems were proposed in literature. In general, HARQ systems are known in literature as ARQ systems with incorporated error correction techniques. A straight forward approach would provide each packet with a fixed error correction code. The underlying idea is to lower the block error probability with the help of the error correction code before the correctness is verified with the error detection code. However, if receiver still discards erroneous decoded versions after an unsuccessfull error correction attempt, these HARQ systems do not really provide a solution. Instead, the congestion problem is simply shifted towards worse channel conditions on the expense of an overall lowered data rate.

To provide an actual solution to this problem, the ARQ system must utilize all previously received transmissions (although erroneously decoded) in addition to the currently received transmission for the decoding process. If done properly, the probability of a decoding failure can steadily be lowered with each retransmission.

Again, in the above presentation of ARQ systems, it was assumed that the ARQ is located in layer 2. Yet the author likes to emphasis, that the provision of a reliable service of layer n for layer $n+1$ by no means requires an ARQ protocol in the layer 2. Instead, the ARQ protocol could run

with all the described principles anywhere from layer 2 to layer n. As an example, layer 3 could use an unreliable service of the data link layer and implement its own ARQ protocol to provide a reliable service to the transport layer

However, the problem, which arises when layer n is the first layer in which an ARQ protocol is implemented is that a single error in the physical layer results in a retransmission of a complete layer n PDU. Usually, the size of the PDUs increases with the hierarchy of the layers and hence the delay as well as the overhead introduced by the ARQ system is larger than it would be for a layer 2 ARQ protocol. Also, each new layer introduces new sources of errors. Ideally, for a reliable communication between to peers in layer n, this layer as well as all layers below should use reliable communication with their corresponding peer entities. The ARQ protocols in the individual layers should only address the sources of errors which are specific to this layer. A violation of this principle may result in an additional performance degradation. As an illustration, consider the reliable transport layer protocol TCP. This protocol was designed to handle the main source of errors in this layer: packet losses due to flooding of a slow network component such as a router. As an adequate response action, the TCP protocol effectively slows down its transmission rate to disburden the overstrained router. If, on the other hand, a TCP/IP packet is lost due to a physical layer error, the best behavior would be to retransmit the erroneous packet as soon as possible without lowering the TCP transmission rate. Yet, the TCP protocol follows its algorithm, leading to a zero TCP throughput for highly erroneous physical channels. Therefore, for optimization reasons, an ARQ protocol should run in the lowest possible layer allowing for such a protocol, i.e. the data link layer, to directly take care of physical layer errors. Besides these arguments, also the realm of possibilities, which a combination of layer 1 and 2 offers, accounts for the implementation of an ARQ protocol in layer 2. This issue will be discussed in the following section.

1.4 Hybrid Layer 1/2 Schemes

The layer 1 of the OSI model is concerned with the (unreliable) transmission raw data. The input as well as the output are layer 2 PDUs. As mentioned, the actual realization of the transmission process is, like the underlaying transmission medium, of no concern to the higher layers.

Again, the layer 1 PDU is not necessarily a bit. If, for example, a M-level modulation is used, the smallest transmitted unit is a modulation symbol which carries M bits. Also, as already mentioned, if transmission codes such as softdecision decoded convolutional codes are used, no single bit can be obtained before the decoding procedure and hence the layer 1 PDU is a codeword of that code.

On the other hand, if one insists that all error correction schemes are part of layer 2, then there is no independent layer 1 in the sense of the OSI model. Instead, layer 1 and 2 merge into what we will call a hybrid layer 1/2.

A hybrid layer 1/2 performs the task of the two lowest layers of the OSI model in a single unit in which no direct boundary can be drawn. Hence, the task of the hybrid layer 1/2 is to

transmit information over a physical channel with a well defined QoSs. Although these hybrid layer violate the principle of the OSI model, the performance gain which can be achieved stands for such an implementation. In addition, the main idea of the separation between the individual layers is to simplify the development process. In practice, however, layer 1 and 2 are due to their close relation anyway developed by the same teams. Hence, a unified treatment of this two layer does not really provide a severe contradiction with the main idea of the OSI model.

Besides the transmission codes there are other task of layer 1 whose performance can be improved if done with the help of the correction codes. Examples are code assisted symbol and carrier synchronization or iterative channel estimation via the embedded code.

In a layer 1/2 hybrid schemes, of course, not all tasks have to be done in a hybrid way. As an example, the the channel estimation can still be done in the old fashioned way although the forward error control coding is implemented in a hybrid way. As another example, an ARQ protocol can run independently on top of hybrid layer 1/2 coded transmission scheme using Turbo codes. In fact, the majority of ARQ systems treated in literature are constructed this way and most ARQ papers simply deal with the integration of new error correction codes into this principle.

However, although such a separation of the ARQ and the (coded) transmission is possible, we expect a performance loss compared to a fully combined system. With this in mind, we can refine the main objectives of this work made on Page 4:

> Principles of efficiently providing reliable services for the data link layer of wireless communication systems via ARQ systems incorporated in hybrid layer 1/2 and their performance analysis are the main topics of this work.

In order to achieve this, Chapter 2 sets up a framework of terminology for ARQ systems and resolves the contradictions encountered in this context in literature. Since the main emphasis is on wireless channels, Chapter 3 introduces wireless channel models such as the additive white Gaussian noise (AWGN) channel and the Rayleigh channel, but also wireless channels with a temporal progression used in the UMTS standardization process. If systems are to be compared, this must be done on the basis of well defined performance measurements. Except for the throughput, for which not even a commonly agreed on definition exist, hardly any other performance measurements were introduced in literature. Therefore, Chapter 4 quite elaborately introduces, defines, and discusses performance measurements for ARQ systems and reveals their mutual relation. Then, Chapter 5 shows some known methods of how the physical layer and ARQ methods can be combined for performance improvements and presents new concepts as another integral part of this work. The performance of the new systems are compared with theoretical bounds and standard implementations. Finally, Chapter 6 concludes the work with a summary and a perspective for future research.

Chapter 2

Classification of ARQ Systems

The aim of this chapter is to provide a common basis of terminology for ARQ systems discussed or presented in this work. Moreover, classification based on the retransmission process is used inconsistently throughout the literature and Section 2.3 shall therefore provide a suggestive basis for future classifications of ARQ schemes.

2.1 Classification by the Transmission Protocol

As we pointed out in Chapter 1, the basic principle of Automatic-Repeat-Request (ARQ) systems is that each received packet is acknowledged. Depending on the decoding process, a positive (ACK) or a negative (NAK) acknowledgment is sent to the transmitter. The transmission protocol, which controls the temporal sequence of packet transmissions in dependency of the acknowledgments of previously sent packets, can be used to classify ARQ systems. Three basic schemes are distinguished:

- Stop-and-Wait
- Go-Back-N
- Selective-Repeat

A system, which sends a new packet only after the previous packet is acknowledged with an ACK, is called a **Stop-and-Wait** (SW) ARQ system [Lin83]. Figure 2.1 illustrates the temporal sequence of packet transmissions based on the received acknowledgments: After sending a packet, the transmitter remains idle until the acknowledgment of that specific packet is received. In case of an ACK, the succeeding packet is transmitted and in case of a NAK, the packet is retransmitted[1].

[1] In the following we will see that more advanced ARQ systems do not necessarily repeat the erroneously decoded packets. Instead some additional information might be sent. For the aim of classification by the transmission process this is of no importance. However, the reader should keep in mind that the terminology 'retransmission of a packet' is used throughout this work for the transmission of any additional information aiding the decoding process of the erroneous packet. Repetition of the packet represents just a special form of additional information.

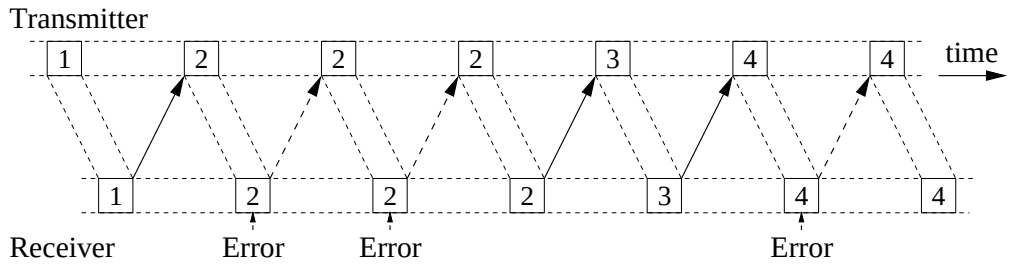

Figure 2.1: Temporal Sequence of Transmissions for a Stop-and-Wait ARQ System. Solid arrows represent ACKs whereas dashed arrows stand for NAKs.

The advantage of the Stop-and-Wait transmission principle is its simplicity. Only a buffer for the currently transmitted packet and the currently received packet is needed on the transmitter and the receiver, respectively. Also, the sequence of the packets is inherently preserved and therefore no packet numbering is required. On the other hand, the simplicity has its price, namely its inherent inefficiency: The next packet can be sent only after the acknowledgment has been received. In some environments, this delay is a multiple of the packet duration, resulting in large idle times of the transmitter, which heavily reduce the effective data rate compared to FEC systems with the same symbol rate.

Although this and related issues will be discussed in detail in Section 4.5 and 4.7, a short example shall serve as illustration. Therefore we consider Digital Video Broadcasting (DVB) over a geostationary satellite (distance from Frankfurt, Germany $\approx$ 37600 km). The round trip propagation time of the signals (packet and ACK/NAK) is roughly

$$t_{prp} = \frac{4 \cdot h}{c} = \frac{4 \cdot 37600\,km}{3 \cdot 10^8\,\frac{m}{sec}} = 0.5\,sec.$$

The highest symbol rate of DVB, specified in [DVB97] is $42.2 \cdot 10^6$ QPSK symbols per second. Accordingly, the packet length of 204 *byte* results in a packet duration of $19.3\,\mu sec$. Hence, the idle time is about 26000 times the packet duration and the transmitter remains at least in $99\,996\%$ of the time in idle.

Besides its inefficiency for environments with a large round trip delay, there is also a practical aspect which limits the use of a SW protocol. Usually, the length of information packets and acknowledgments differ considerably. In two-way communication systems, where both sides can be transmitter and receiver, different protocols must be used, depending whether a message or an acknowledgment was sent. It would be preferable to collect a certain amount of acknowledgments until the information packet size is reached, but due to the Stop-and-Wait principle this is not feasible. Additionally, in wireless applications each packet must be preceded by a header (detection, synchronization, etc.). Acknowledgments are very short messages, but still need a header and hence the usage of the back channel is larger than it would be with the collected acknowledgments.

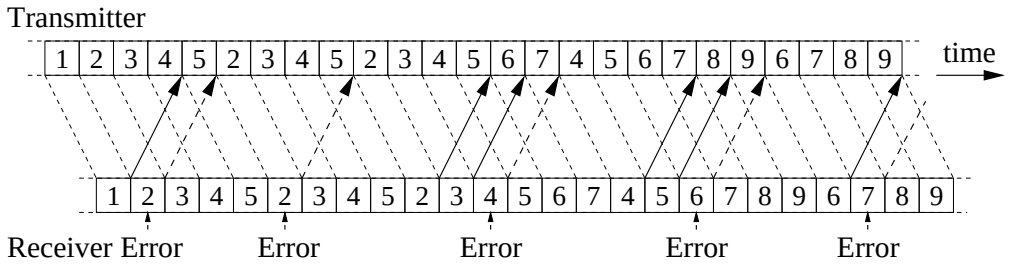

Figure 2.2: Temporal Sequence of Transmissions for a Go-Back-N ARQ System. Solid arrows represent ACKs whereas dashed arrows stand for NAKs.

An improvement of ARQ systems can be achieved by letting the transmitter send packets continuously, i.e. succeeding packets are sent before the preceeding packets are acknowledged. As a consequence, the *transmitter* has to buffer sent packets for retransmission until they are positively acknowledged. Therefore, a memory at the transmitter side is needed (see Figure 1.3 on Page 8). If a NAK is received, the continuous transmission is stopped and the negatively acknowledged packet is retransmitted. The question which then arises is what packet is sent next. If no *receiver* memory is available the only choice is to continue with the temporal successor of the retransmitted packet. That is, if a NAK for packet N is received, the transmission is stopped, packet N is retransmitted and the transmission is continued with packet $N + 1$, although it has already been transmitted. A system with this transmitter behavior is called a **Go-Back-N** (GBN) ARQ system (see Figure 2.2). If the decoding of a packet failed, a NAK is sent back and the transmitter retransmits the erroneously decoded packet and precedes with the succeeding packets since they were not processed by the receiver. All packets sent after the first transmission of the erroneous packet and the arrival of its negative acknowledgment are lost (in the above example 26 packets).

Clearly, the effective data rate of an ARQ system with GBN protocol is always superior to that of the corresponding system with SW protocol. Specifically, if no errors occur, each packet is positively acknowledged and delivered to the sink. Due to the continuous transmission, the effective data rate is identical to that of the corresponding FEC system. On the other hand, if due to the channel condition the first transmission of a packet fails with a probability approaching one, the effective data rates of a Stop-and-Wait and a Go-Back-N system merge. A detailed performance analysis of the GBN protocol will be presented in Section 4.5 and 4.7. As mentioned, the GBN protocol requires a transmitter memory. Section 4.6.1 investigates these requirements in more detail.

In a Stop-and-Wait system, there are no ambiguities about the identity of the received packet, since each packet is transmitted one after another. In a Go-Back-N system after sending a NAK, however, a received packet is either the requested retransmission or one of the $(N - 1)$ others packets send after the first transmission of the erroneous packet. Especially if the roundtrip delay is variable, the transmitter must somehow mark (e.g. short transmission gap, or special header) the retransmission. This little complication, however, is rewarded with the possibility of collective

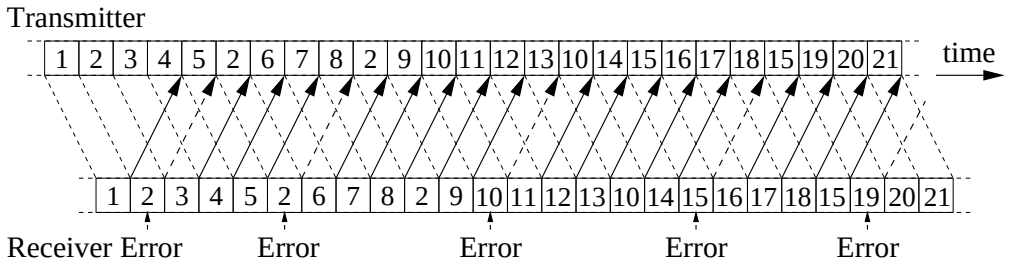

Figure 2.3: Temporal Sequence of Transmissions for a Selective Repeat ARQ System. Solid arrows represent ACKs whereas dashed arrows stand for NAKs.

acknowledgments[2] (as range-acknowledgment for example). Hence, the back-channel usage can be reduced in comparison to systems with SW protocol. However, as soon as the decoding of a packet fails, a NAK must be sent to keep the transmission process going. Therefore, no multiple negative acknowledgments can be collected.

The continuous transmission of a Go-Back-N system improves the effective data rate considerably compared to a Stop-And-Wait system, especially for good channel conditions and large round trip delays. Still, the transmitted packets following a packet detected in error are lost, even if they would have been decoded correctly. With an additional receiver buffer (see Figure 1.3) these packets could be processed and saved in case of a successfull decoding. By this, only erroneous received packets are required to be retransmitted. ARQ systems of this type of protocol are called **Selective-Repeat** (SR) ARQ systems.

Figure 2.3 shows the temporal sequence of sending packet depending on the acknowledgments. Having no transmission gaps and no unnecessary retransmissions, the Selective-Repeat is the most efficient of the three discussed protocols. Besides the mentioned additional receiver buffer (Section 4.6.2 discusses this issue), reordering of the successfull decoded packets becomes necessary if the correct order shall be preserved. The Selective-Repeat protocol additionally allows for multiple negative acknowledgments.

Additionally to the discussed basic transmission schemes, mixtures of these are possible. If, for example, the receiver buffer of a Selective-Repeat ARQ is exceeded, the system has to switch to a form of Go-Back-N mode.

2.2 Literature Review

In Chapter 1, the principle of the ARQ strategy was presented. These feedback systems exist in theory as well as in practice for several centuries. In the following section we briefly investigate the origin and the historical development of the terminology ARQ and then look in Section 2.2.2 into further classifications of these systems as found in literature. This section will reveal inconsistencies in the literature, as well as argues that the applied way of classification is not

[2] Multiple acknowledgments have impact on the transmitter memory requirements. This issue is not analyzed in Section 4.6.1.

meaningful. Therefore, Section 2.3 seizes suggestions for classification found in early papers of feedback communication systems and further develops this classifications to a meaningful frame of classification for all existing ARQ systems. This classification will then be used throughout the remainder of this work.

2.2.1 Historical Review

Years before Shannon wrote his famous paper in 1948, the first practical feedback systems were already in operation. One of these first system was a long-range teletypewriter service, called by its inventor Van Duuren, *the* ARQ (automatic request for repeat) system [Schw63]. In the following two decades after Shannon's paper, researchers became more and more interested in general communication systems with feedback channels. The community spoke from information feedback systems [Cha56], retransmission error control systems [Kuh63], sequential transmission schemes with feedback [Hor63], and retransmission systems [Ben64], while for the mentioned practical system the terminology 3-out-of-7 ARQ was used [Nes63]. In the 1970's some authors (Morris [Mor78][Mor79], Sastry [Sas75]) started to use ARQ as a general name for feedback systems, while others used a more descriptive terminology to name their feedback systems, among them Mandelbaum in 1974 [Man74] and Sindhu in 1977 [Sin77].

In the 1980's, the majority of the coding community collectively used the terminology ARQ for general feedback systems. The publication of the Lin's and Costello's standard text book [Lin83], who devoted an own chapter to error control with feedback system, made a significant contribution to this trend, since they also called these systems ARQ systems and may authors started to reference this book.

2.2.2 Classification in Literature

Throughout the years, many attempts were made to improve the performance of the pure ARQ principle of error detection and retransmission. A key paper in this sense, represents the 1970 paper of Rocher and Pickholtz [Roc70]. They presented a what they call Hybrid Transmission Scheme, which utilizes an error correction in addition to an error correction code. Purpose of the error correction code is to lower the retransmission probability for noisy channel conditions. A conceptual similar system with was given by Sastry [Sas76] and ARQ systems which provide error correction capability in the transmission became known as **Hybrid ARQ** (HARQ) schemes. This terminology is still in use today.

Another conceptual key paper appeared in 1974 by Mandelbaum [Man74]. In this paper he presented a feedback coding scheme where the redundancy is delivered in small increments and all transmissions are required for decoding. This scheme differs from all existing systems in two aspects. Firstly, the retransmissions are not simply repetitions, but are used to construct sequentially more and more powerful codes. And secondly, the scheme requires receiver memory. Another scheme, which also makes use of several retransmissions for the decoding process was described by Sindhu in 1977 [Sin77]. Contrary, to the incremental redundancy scheme of Man-

delbaum, Sindhu's system, however, used simple repetitions. As a result, the decoder not necessarily requires all previous transmissions for a decoding attempt. In 1979, Metzner's presented a retransmission scheme [Met79], which makes use of several transmissions (as Mandelbaum's and Sindhu's systems), but the retransmissions are not simple repetitions (contrary to Sindhu's system) and all retransmissions can be decoded independently (contrary to Mandelbaum's scheme). By 1979, all basic conceptual systems were already presented, however the inventors of these schemes did not suggest any special terminology for their systems.

The classification mess, as we see it today, started with the paper of Lin and Yu in 1982 [LiYu82]. They presented an ARQ scheme which is conceptually identical to the mentioned system of Metzner and they called the class of systems, where redundancy (*"parity check bits for error correction"* as they call it) is delivered in increments, a Type-II Hybrid ARQ system, whereas Hybrid ARQ systems of Type-I are the ones with constant error correction capability in each transmission. This represents a wrong concept of classification, since each retransmission represents additional redundancy and therefore all ARQ system are of Type-II. Hence it depends on the decoder realization if an ARQ system belongs to the Type-I or Type-II class. To illustrate this consider Metzner's system. As said, each retransmission is self-decodable, i.e. the information can be extracted from each transmission. A cheap receiver realization could save the memory, which is required for the described decoding procedure on the expense of performance degradation. This realization would be a Type-I Hybrid ARQ system. A realization with memory, however, would be a Type-II HARQ system.

To make things even worse, the Type-I/II HARQ classification made its way into the mentioned standard text book [Lin83] wherein Lin called their *specific* realization in [LiYu82] as *the* Type-II HARQ system. Luckily, most authors called all HARQ systems with the above mentioned principle of memory as Type-II systems.

So up to the year 1995, the world a ARQ systems is classified in literature into 3 classes: Pure ARQ systems, Type-I HARQ systems, and Type-II HARQ systems. This classification is illustrated in Figure 2.4.

In 1995, however, Kallel's paper [Kal95] was published, wherein he introduced a new classification type, yet without introducing a conceptual new ARQ scheme. More specific, he presented an ARQ with incremental redundancy, where the retransmissions are not necessarily repetitions, but are self-decodable and called these ARQ schemes Type-III HARQ schemes. Contrary, ARQ schemes with non-selfdecodable retransmissions he referred to as Type-II HARQ schemes. This definition yields to a total mess. Lin and Yu's famous Type-II scheme [LiYu82] has selfdecodable retransmissions and is according to Kallel's definition from now on a Type-III scheme, and the same applies to Metzner scheme from 1979 [Met79]. In fact, the majority of up to 1994's Type-II systems have selfdecoding retransmissions and are now transformed into Type-III schemes. In the group of Type-II HARQ schemes, according to Kallel's classification, only a few remainders are left, among these Mandelbaum's 1974 system [Man74], Dorsch's adaptive forward error correction for channel with feedback [Dor86], and Hagenauer's rate-compatible punctured convolutional code scheme [Hag88]. As a result from this definition, if one refers to a Type-II HARQ scheme, it must be accompanied by a year specification! Now, one should believe, that this an-

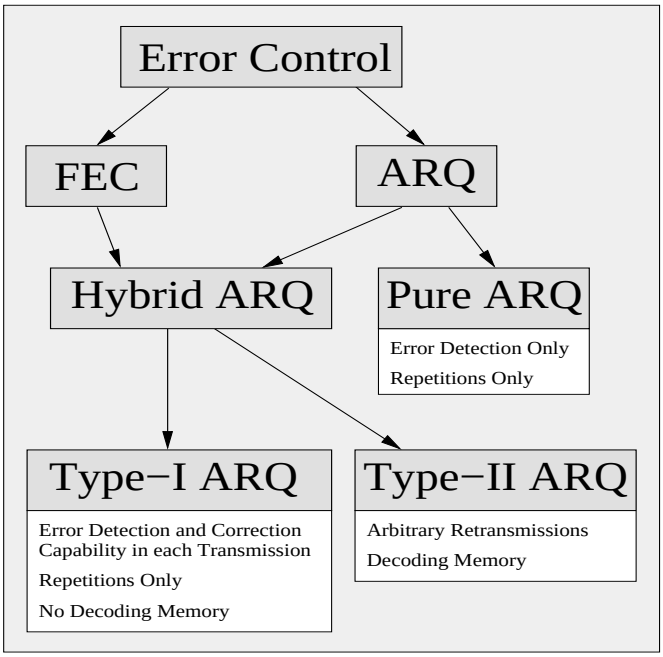

Figure 2.4: Classification of Error Control Systems up to the Year 1994.

noying definition never made it beyond Kallel's paper. Unfortunately, nothing is further from the truth. In fact, it made its way into the specification of one of the most discussed communication systems today, namely UMTS.

2.3 Classification by the Encoding Protocol

In the literature review we saw that in the present ARQ classification mainly two things got mixed up: namely, what is retransmitted and how does the encoder make use of the transmitted redundancy, i.e. how does it decode the sequence of transmissions. The way the decoder uses the transmitted information, however, may not be used for a classification of the ARQ system since it is simply a decoder implementation issue. In the same way, FEC schemes are not classified by decoder implementation. There are no, let's call it Type-I/II convolutional codes, where the first group represents hard-decision decoded convolutional codes and the second group a softdecision decoded convolutional codes.

> An ARQ system classification must be independent of the decoder strategy and purely be based on the algorithm of generating new transmissions and retransmissions.

In addition, for the sake of clearness, a classification should essentially be self-explanatory and no generic definitions like Type-I/II/III should therefore be used.

In Section 2.1 we classified the ARQ system according to the *temporal* sequence of the transmissions (original transmissions and retransmissions), i.e. *when* a new packet or a retransmission

is sent. This classification was, as postulated, self-explanatory and not related to any decoder implementation.

In this section we will present a classification according to the encoder protocol, i.e. *what* is sent at the $k-th$ transmission, independent of the fact, *when* it will be sent.

If we consider general systems with forward and backward channels, we can distinguish *two-way communication systems* and *feedback systems*. The latter one uses the backward channel only to provide the transmitter some kind of feedback about the status of the transmission process and the actual information flow is unidirectional. Two-way communication systems, on the other hand, transmit information in both ways. Of course, the two way communication can be used to implement two feedback systems by appending the feedback information on the information packets in the reverse channel. This distinction between this two system groups with forward and backward channel was already presented by Schwartz in 1963 [Schw63]. The concept of attaching feedback to the information packets in the reverse channel is today know as piggybacking and was first presented conceptually by Chang in 1961 [Cha61].

The class of feedback system can again be subdivided according to the information which is returned. We want to adapt Schwartz's distinction [Schw63] between *decision* feedback and *information* feedback systems. Decision feedback simply report the decision *of the receiver* about the acceptance of the transmitted information to the sender, i.e. a binary information such as a ACK or a NAK is returned. Information feedback systems, on the other hand, return more detailed information than a binary decision variable back to the transmitter. This might be some information about the received message and, as described in [Cha56] and [Schw63]. The decision is up to the *transmitter*, or still a decision of the *receiver* is reported but with some form of additional reliability information. This separation is depicted in Figure 2.5. Actually, Shannon depicted in his famous 1948 paper [Sha48] a perfect (unfortunately imaginary) information feedback system (see Figure 4.14 on Page 74) and based his explanations of the channel capacity on this system. Section 4.4.2 briefly discusses his results.

By now, we have set up the required terminology to state a definition for ARQ schemes which will be used throughout the remainder of this work:

> Feedback systems with decision feedback will be called automatic retransmission request (ARQ) schemes.

The author purposely avoids to call these schemes automatic repeat request schemes, since it possibly awakes the impression that the retransmission are necessarily repetitions. This, however, is only true for a specific subclass of ARQ schemes which will be denoted as the constant encoder (CE) subclass of the ARQ systems:

> ARQ systems which send repetitions of negatively acknowledged transmissions as retransmissions are called constant encoder ARQ (CE-ARQ) systems.

The remaining ARQ systems, i.e. the systems where at least some retransmissions are no repetitions of the original message, will be called variable encoder (VE) ARQ systems (due to the varying encoder output):

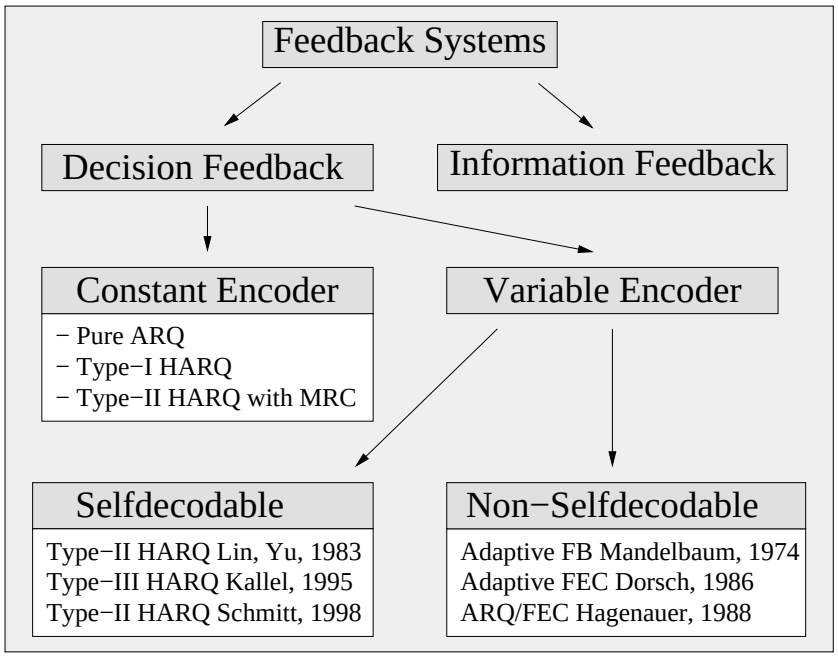

Figure 2.5: New Classification of Feedback Systems according to Their Encoding Protocol.

> ARQ systems which send at least some retransmissions which are not identical with the original message are called varying encoder ARQ (VE-ARQ) systems.

As Figure 2.5 already depicts, Pure ARQ, Type-I HARQ and Type-II HARQ with repetitions and diversity combining methods such as code combining [Chs85] [Kal90] and maximum-ratio-combining [Bre59] [Har94] belong to the CE-ARQ subclass.

The class of VE-ARQ systems again is subdivided into systems where all retransmission are self-decodable, i.e. the information can be extracted from *any* retransmission alone and system where at least some retransmissions are not self-decodable. Most systems belong principally to one or the other group but there are also some systems, such as the one in [Hag88], where it depends on the actual realization, i.e. the amount of information sent in the retransmission, if the system has self or nonself decodable retransmissions.

2.4 Classification by the Decoding Procedure

The classical way of classification did not distinguish between system properties and decoder realizations. In the preceding section, we therefore introduced a classification only based on the encoding process. This classification, as well as the classification based on the temporal sequence (SW, GBN, SR) represent system classifications.

In this section, we add another classification, this time based on the actual decoder realization. The most basic distinction, if it comes to decoder realization, is if the encoder bases its decoding

attempt solely on the currently received transmission or uses two or more previously received transmissions for decoding. In order to enable the latter one, the receiver must have a memory for saving the required information for the already received transmissions. As mentioned in Section 2.2.1, the first systems which made use of several (actually all) received transmissions was Mandelbaum's incremental redundancy scheme [Man74]. Another conceptual similar scheme was presented by Sindhu in 1977 [Sin77] and he specifically mentioned the fact that the system requires receiver memory. Moreover, he introduced a new system name:

> "*An ARQ-With-Memory (MRQ) system ... is an ARQ system that is provided with additional logic and memory at the receiver for the purpose of correcting errors using retransmitted blocks*" .

Although this definition did not prevail in literature with the exemption of a few authors, among them Metzner [Met79] and Benelli [Bnl87] [Bnl92], we want to seize his definition. Following this definition, but keeping in mind that the decoding realization should not be part of a system classification we define a classification of the decoding procedure:

> The decoding procedure in an ARQ system which makes jointly use of two or more transmissions is called memory decoding procedure and an ARQ system employing this strategy is called a memory ARQ (MARQ).

Again, the author likes to emphasis that the term MARQ denotes a specific realization of an ARQ system, i.e. for the same ARQ system there are various decoder realizations with and without memory for the decoding process. An exception are VE-ARQ systems with non-selfdecodable retransmission. Obviously, for these systems memory is inherently required.

Chapter 3

Channel Models

The purpose of a digital communication system is the transmission of digital information from a source (transmitter) over a certain medium (channel) to the destination (sink). Digital information means that the message to be transmitted is composed of a sequence of discrete numbers. For transmission, these numbers have to be mapped onto analoge signals, suitable for transmission over the channel. During the transmission and reconstruction process, many kinds of disturbances can afflict the communication link. Examples include:

- Attenuation
- Fading (time-varying attenuation)
- Additive noise (thermal noise from sky and lossy circuits, impulse noise from lightning and ignition systems, ...)
- Bandwidth limitations (bandlimited transmission medium, forced bandwidth limitation through filters)
- Interference (co-channel or intracell interference, jamming, crosstalk, ...)
- Delays

Unfortunately, a channel model that encompasses all the impairment above is complex and becomes intractable to analyze. The different channel models used in this work take only some of the above mentioned disturbances into account. Before we present the individual channel models used throughout this work we describe the general transmission process to set up a common terminology.

3.1 Generic Transmission Process

Figure 3.1 depicts a block diagram of the transmission process. The information bit sequence $\underline{b} = (b[0], \ldots, b[k], \ldots)$ is digitaly processed (e.g. source and channel coded) to form the bit se-

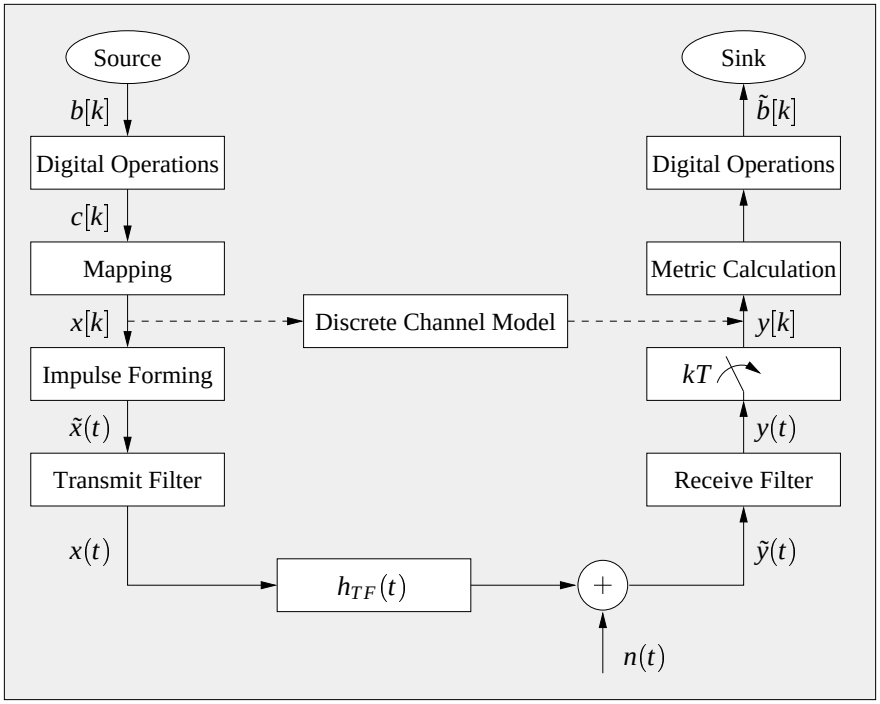

Figure 3.1: Block Diagram of Digital Transmission Scheme.

quence $\underline{c} = (c[0], \ldots, c[k], \ldots)$ to be transmitted. The mapper transforms this bit sequence into the transmit sequence $\underline{x} = (x[0], \ldots, x[k], \ldots)$ by mapping a certain number of bits onto so-called channel symbols, taking one of M distinct values $x_0, \cdots, x_{M-1}$. These discrete symbols correspond to M analog signals. In general, this could be M totally different signals, but practically they are linear combinations of a set of basis functions. The number of basis functions is the dimensionality of the signal set, for example one in binary modulation schemes or M-ary ASK and two for M-ary PSK or M-ary QAM with $M > 2$. Since the dimensionality of practical modulation schemes is 2 at maximum, the linear combination can be nicely represented by complex numbers, i.e. $x_i \in \mathbf{C}$, $i = 0, \ldots, M-1$. The conversion process of the complex transmit sequence $\underline{x}$ into complex analoge signals can be mathematically presented via the modulation of an impulse train with symbol rate $\frac{1}{T}$ with this transmit sequence

$$\widetilde{x}(t) = \sum_{k=0}^{\infty} x[k] \cdot \delta(t - k \cdot T)$$

and filtering with the transmit filter (pulse shaping) with impulse response $h_{TF}(t)$, thereby generating the analoge signal

$$x(t) = \sum_{k=0}^{\infty} x[k] \cdot h_{TF}(t - k \cdot T).$$

Therefore, the analoge transmit signal is composed of a temporal sequence of the M possible

waveforms. Each individual signal $x[k] \cdot h_{TF}(t - k \cdot T)$ has an energy (transmit energy) of[1]

$$E_S^{(Tx)}[k] = |x[k]|^2 \cdot \int_0^{T_S} h_{TF}(t)^2 \, dt.$$

Without loss of generality, the gain of the transmit filter can be assumed to be one, so that a complex channel symbol with amplitude $|x[k]|$ results in a transmit signal with energy $E_S^{(Tx)}[k] = |x[k]|^2$. If the assumption can be made that all transmit symbols x_i are equally likely, the average transmit energy $\overline{E_S^{(Tx)}}$ is

$$\overline{E_S^{(Tx)}} = \frac{1}{M} \cdot \sum_{k=0}^{M-1} |x_k|^2 . \tag{3.1}$$

For PSK, where all transmit symbols have the same amplitude the average transmit energy is equal to the individual symbol energies

$$\begin{aligned} \overline{E_S^{(Tx)}} &= E_S^{(Tx)} \\ &= |x_k|^2 . \end{aligned}$$

In higher level ASK and QAM modulation, however, the average transmit energy is a function of the amplitude of the individual transmit symbols. A detailed treatment on the relation of the average transmit energy to the complex coordinates for these modulation schemes can be found in [Kre89].

The waveforms are transmitted over a possibly time varying frequency selective channel[2], characterized by its impulse response $h_{TM}(t, \tau)$. This part of the model describes the actual distortion of the transmitted signal. In the models used throughout this work, this part is either a constant attenuation (in the case of AWGN), or a multiplicative time varying factor $\mathbf{r}$ (Rayleigh channel and the wide sense stationary uncorrelated scattering (WSSUS) model). The distorted signal is then additionally corrupted by additive noise $\mathbf{n}(t)$. The received signal $\widetilde{\mathbf{y}}(t)$ is passed to the receive filter (matched filter) $H_{RF}(f)$ and sampled at a rate equal to the symbol rate, yielding the receive symbols $\mathbf{y}[k]$. Due to the bandlimitation of the filters and the channel, the transmission system might suffer from intersymbol interference (ISI). For its prevention the product

$$H_{TF}(f) \cdot H_{TM}(f) \cdot H_{RF}(f)$$

of the filter and channel frequency responses must satisfy the Nyquist pulse-shaping criterion [Pro95]. Then, the received signal is

$$\mathbf{y}(t) = \sum_{k=0}^{\infty} \mathbf{r}[k] \cdot x[k] \cdot x(t - kT) + \widetilde{\mathbf{n}}(t)$$

[1] The symbol duration T_S is not necessarily equal to the inverse of the symbol rate $\frac{1}{T}$.

[2] The optional modulation of the baseband signal $x(t)$ into the transmit frequency is omitted. Instead the equivalent baseband model of the channel is used.

with $\mathbf{r}[k]$ being a complex attenuation and phase distortion factor[3] and

$$\widetilde{\mathbf{n}}(t) = \int_{-\infty}^{\infty} \mathbf{n}(\tau)\, h_{RF}(t-\tau)\, d\tau.$$

Since the transmission line is assumed to satisfy the Nyquist criterion, $x[k] \cdot x(t-kT)$ is zero at multiples of the symbol time T except for $t = 0$ and consequently

$$\mathbf{y}[k] = \mathbf{y}(kT) = \mathbf{r}[k] \cdot x[k] + \widetilde{\mathbf{n}}[k]. \tag{3.2}$$

Hence, the samples $\mathbf{y}[k]$ are the amplified and phase shifted versions of the transmitted symbols $x[k]$ with a superimposed noise part. Equation 3.2 describes the equivalent discrete channel model (see Figure 3.1) and for simulation purposes, the effect of modulation, analoge channel, and demodulation can be simulated by multiplying the complex channel symbol $x[k]$ with the factor $\mathbf{r}[k]$ and adding the noise sample sequence $\widetilde{\mathbf{n}}[k]$. Consequently the channel models focus on the generation of the attenuation factors according to a specific distribution and temporal correlation and the generation of the noise. For the noise sequence, however, it must be observed, that the receive filter transforms the autocorrelation $s_{\mathbf{nn}}(\tau)$ of the channel noise. If $S_{\mathbf{nn}}(f)$ is the noise power density, the filtered noise has a power density of

$$S_{\widetilde{\mathbf{nn}}}(f) = |H_{RF}(f)|^2 \cdot S_{\mathbf{nn}}(f) \tag{3.3}$$

Hence, if also $|H_{RF}(f)|^2 \cdot S_{\mathbf{nn}}(f)$ satisfies the Nyquist pulse-shaping criterion, the autocorrelation $s_{\widetilde{\mathbf{nn}}}(\tau)$ is zero at the sampling instants and the samples $\widetilde{\mathbf{n}}(kT)$ are independent.

In the following sections, different models with increasing degree of complexity are presented.

3.2 Additive White Gaussian Noise Channel

The Additive White Gaussian Noise (AWGN) channel is the simplest continuous channel model.

Additive in this context means, that there is **only** an additive noise part and the attenuation $\mathbf{r}[k]$ is limited to a real constant r for all symbols. Hence, the deterministic part of the sampled receiver output is $r \cdot x[k]$ and the corresponding receive energy is

$$E_S^{(Rx)}[k] = r^2 \cdot |x[k]|^2,$$

and the average receive energy is

$$\overline{E_S^{(Rx)}} = r^2 \cdot \overline{E_S^{(Tx)}}.$$

Without loss of generality, the attenuation r can be set to 1 in order to avoid the distinction between the transmit and the receive energy.

[3] If the variations of the channel are assumed to be small compared to symbol duration T, it can be represented as a constant factor for that symbol.

White refers to the constant spectral characteristics of the noise $\mathbf{n}(t)$ depicted in Figure 3.1. As a consequence all noise samples $\mathbf{n}[k]$ are mutually statistically independent and the channel therefore memoryless.

Gaussian, finally, refers to the distribution of these noise samples, namely a Gaussian or normal distribution. The reason why this distribution was chosen for a generic model can be found in the central limit theorem of probability theory and in the sources of noise. Main sources of additive noise are receiver noise resulting from thermal agitation of the electric charge carriers in resistive materials, shot noise from diodes and semiconductor devices, atmospheric, and galactic noise [Yac93]. All of these forms of noise have sources which are immense in number, do not have any preferred direction, and are constant in intensity. As a result, the unfiltered random signal $\mathbf{n}(t)$ can be modeled by a stationary noise process, which amplitudes obey a constant mean free normal distribution. Its temporal progression is statistically described by its autocorrelation function $s_{\mathbf{nn}}(\tau)$ or its power density $S_{\mathbf{nn}}(f)$ which is, as above argued, constant. If the constant power spectrum density of the unfiltered noise sequence is $\frac{N_0}{2}$ (in $\frac{W}{Hz} = J$) for positive and negative frequencies, the autocorrelation $s_{\mathbf{nn}}(\tau)$ (ACF) and its Fourier transform, the power spectrum density $S_{\mathbf{nn}}(f)$ (PSD), are

$$\begin{aligned} s_{\mathbf{nn}}(\tau) &= \tfrac{N_0}{2}\delta(\tau) \\ S_{\mathbf{nn}}(f) &= \tfrac{N_0}{2}. \end{aligned}$$

According to Equation 3.3, the power spectrum density $S_{\widetilde{\mathbf{nn}}}(f)$ of the filtered noise part is dependent on the correlation function of the impulse response of the matched filter. With $r = 1$ and the already mentioned assumption, that the receive filter satisfies the Nyquist criterion, the samples at the receiver can be represented as

$$\mathbf{y}[k] = x[k] + \widetilde{\mathbf{n}}[k], \tag{3.4}$$

with $\widetilde{\mathbf{n}}[k]$ being a complex random variable (r.v.) with mutually independent mean free normal distributed real and imaginary part with variance $\frac{N_0}{2}$.[4]

Obviously, the performance of a digital communication system is dependent on the energy of the received signal E_S but also on the energy of the noise, i.e. its variance $\sigma^2 = \frac{N_0}{2}$. However, a not that obvious fact is that the performance is only dependent on the ratio of these two parameters, the so-called signal-to-noise ratio (SNR) [Pro95]

$$SNR[k] = 2 \cdot \frac{E_S[k]}{N_0} \tag{3.5}$$

and its averaged version

[4] The actual representation depends on the gain of the receive filter. However, the crucial ratio of the energy of the deterministic part to the noise variance is independent of this gain.

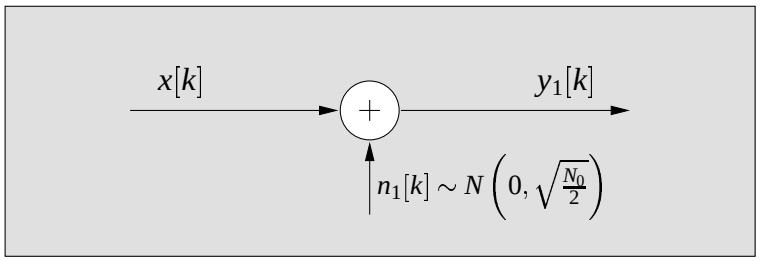

Figure 3.2: Discrete AWGN Model 1. Average Signal Energy E_S.

$$\overline{SNR} = 2 \cdot \frac{\overline{E_S}}{N_0}.$$

Accordingly, any performance measurement of a digital communication system can be plotted versus the (average) signal-to-noise ratio $\overline{SNR}$ or the (average) signal-energy-to-noise-power-density ratio $\frac{\overline{E_S}}{N_0}$. In the remainder of the work $\frac{E_S}{N_0}$ is used although sometimes it is referred to the SNR. The reader is asked to keep in mind the factor 2 between this two ratios.

If an AWGN channel has to be modeled for simulation purposes, the channel model can be reduced to the generation of discrete random variables with the appropriate ratio of the signal energy and noise variance and an addition as given by Equation 3.4. Any combination of modulation symbol energy and noise variance leading to the required $\frac{E_S}{N_0}$ is a valid model. This scheme is depicted in Figure 3.2 and described in the following

The complex signal space representation of the used modulation alphabet $\{x_0, \ldots, x_{M-1}\}$ has an average energy E_S as given by Equation 3.1. Depending on the information to be transmitted, a specific symbol $x[k]$ is chosen. and a complex random variable $\mathbf{n}_1[k]$ is generated via two independent zero mean normal distributed random variables with variance $\sigma_1^2 = \frac{N_0}{2}$. Consequently, the noise samples obey the density function

$$\frac{1}{\sqrt{\pi \cdot N_0}} e^{-\frac{n^2}{N_0}}$$

for each dimension of the signal set. Hence, the complex noise samples $\mathbf{n}_1 = Re\{\mathbf{n}_1\} + j \cdot Im\{\mathbf{n}_1\}$ have the complex density function

$$\begin{aligned} f_{\mathbf{n}_1}(n) &= \tfrac{1}{\pi \cdot N_0} \cdot e^{-\frac{Re\{n\}^2 + Im\{n\}^2}{N_0}} \\ &= \tfrac{1}{\pi \cdot N_0} \cdot e^{-\frac{|n|^2}{N_0}}, \; n \in \mathbf{C}. \end{aligned}$$

The sampled receiver output is then represented by the sum

$$\mathbf{y}_1[k] = x[k] + \mathbf{n}_1[k]$$

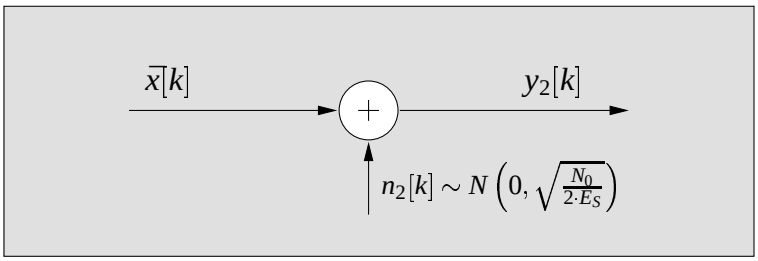

Figure 3.3: Discrete AWGN Model 2. Average Signal Energy 1.

and the conditioned PDF, that the channel output $\mathbf{y}_1 = x + \mathbf{n}_1 = y$ is observed under the condition that $\mathbf{x} = x$ was transmitted is

$$f_{\mathbf{y}_1|\mathbf{x}}(y|x) = \frac{1}{\pi \cdot N_0} \cdot e^{-\frac{1}{N_0} \cdot |y-x|^2}, \; y, x \in \mathbf{C}. \tag{3.6}$$

If a transmission with a different $\frac{E_S}{N_0}$ ratio is to be simulated, either the signal energy, i.e. the modulation alphabet $\{x_0, \ldots, x_{M-1}\}$ is scaled by a real factor, or the variance of the noise is adapted, or even both parameters are changed. Whatever way is chosen is of no importance as long as the proper ratio is obtained and if the decoding is purely based on one receive sequence. Later we will see, that as soon as two transmissions with different $\frac{E_S}{N_0}$ ratios are to be combined (as it will be the case in some ARQ systems), the way this ratios were obtained becomes crucial. This issue will be discussed in detail in Sections 5.1.2 and 5.1.4.

Since the performance is solely dependent on the ratio, it is useful to obtain a channel model with only one degree of freedom, namely this ratio. To do so, either the scaling of the modulation alphabet or the noise variance must be fixed. We decide for the first one and introduce the normalized signal space representation of the modulation alphabet $\overline{x}[k] \in \{\overline{x_0}, \ldots, \overline{x_{M-1}}\}$ derived from the unnormalized alphabet with

$$\overline{x_k} = \frac{x_k}{\sqrt{\frac{1}{M} \cdot \sum_{l=0}^{M-1} |x_l|^2}}.$$

Accordingly, the normalized alphabet has an average energy of

$$\frac{1}{M} \cdot \sum_{k=0}^{M-1} |\overline{x_k}|^2 = 1. \tag{3.7}$$

Again, according to the information one specific symbol $\overline{x}[k]$ is chosen from the normalized alphabet, a complex random variable $\mathbf{n}_2[k]$ is generated, and the sum yields the sampled receiver output $\mathbf{y}_2[k] = \overline{x}[k] + \mathbf{n}_2[k]$ (see Figure 3.3). However, this time the real and imaginary part of $\mathbf{n}_2[k]$ are independent zero mean normal distributed random variables with a variance of $\sigma_2^2 =$

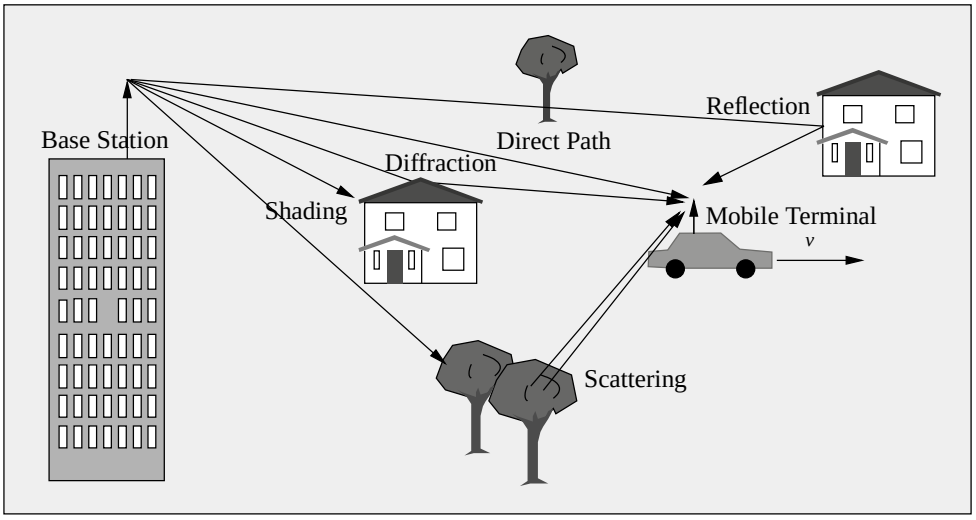

Figure 3.4: Multipath Propagation in an Radio Environment.

$\frac{N_0}{2 \cdot E_S}$, having a complex PDF

$$\begin{aligned} f_{\mathbf{n}_2}(n) &= \frac{E_S}{\pi \cdot N_0} \cdot e^{-\frac{E_S \cdot \left(Re\{n\}^2 + Im\{n\}^2\right)}{N_0}} \\ &= \frac{E_S}{\pi \cdot N_0} \cdot e^{-\frac{E_S}{N_0} \cdot |n|^2}, \; n \in \mathbf{C}. \end{aligned} \tag{3.8}$$

Hence, if the normalized transmit symbol $\overline{x}$ is known, the conditioned PDF of the model output $\mathbf{y}_2 = \overline{x} + \mathbf{n}_2$ is simply the PDF 3.8 with n being substituted by $n = y_2 - \overline{x}$:

$$f_{\mathbf{y}_2|\overline{\mathbf{x}}}(y|\overline{x}) = \frac{E_S}{\pi \cdot N_0} \cdot e^{-\frac{E_S}{N_0} \cdot |y - \overline{x}|^2}, \; y, \overline{x} \in \mathbf{C}. \tag{3.9}$$

3.3 Rayleigh Channel

In the preceding section the signal was corrupted by additive white noise only and the signal itself remained unchanged or was attenuated by a constant value. In a general radio environment, however, the signal experiences multipath propagation as depicted in Figure 3.4, i.e. the transmitted signal reaches the receiver via different main paths such as a direct path, several reflections on different obstacles, or scattered paths. The different main paths may differ considerably in their attenuation and delay. Also, each non-direct path (reflections, diffractions, or scattered paths) consist of a multitude of sub-paths with random phases. As an example consider the reflection of the signal on a rough surface of a building. Contrary to the average delays of the individual main paths, the delays of the paths within a main path will differ only slightly.

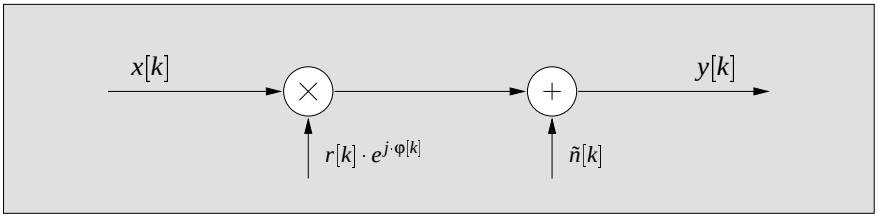

Figure 3.5: Discrete Rayleigh Channel Model Prior To Phase Correction.

3.3.1 The Rayleigh Distribution

The Rayleigh channel models the physical scenario, that only **one non-direct main path** is present. Under this assumption, the received signal is composed of a large number of paths, which have a uniform distribution of the phases (as a result of the small differences in delay), and random attenuation. In [Yac93] it is shown that the resulting signal $\mathbf{y}(t)$ is given by the product of the transmitted signal $x(t)$ with a complex noise process $\mathbf{r}(t) \cdot e^{j \cdot \varphi(t)}$. Herein, the real noise process $\mathbf{r}(t)$ obeys a Rayleigh distribution (hence the name of this channel) given by the probability density function (PDF) [Bro95]

$$f_{\mathbf{r}}(r) = \frac{r}{\sigma^2} \cdot e^{-\frac{r^2}{2\sigma^2}} \cdot u(r) \tag{3.10}$$

and $\varphi(t)$ is independently from $\mathbf{r}$ uniformly distributed with PDF

$$f_{\varphi}(\varphi) = \frac{1}{2\pi} \cdot (u(\varphi) - u(\varphi - 2\pi))$$

with $u(x)$ denoting the unit step function

$$u(x) = \begin{cases} 1 & , x \geq 0 \\ 0 & , else \end{cases} .$$

Therefore, the Rayleigh channel has two additional effects compared to the AWGN channel. Firstly, it randomly attenuates or amplifies the signal and secondly it adds an independent random phase shift. At the receiver, again AWGN $\widetilde{\mathbf{n}}(t)$ with a double sided noise power density of $\frac{N_0}{2}$ is added. Under the assumption, that the continuous Rayleigh noise process hardly changes its statistics during the duration of a symbol, $\mathbf{r}$ and φ can be treated as constants and again a discrete channel model can be obtained on a symbol basis (see Figure 3.5), with $\mathbf{r}[k]$, $\varphi[k]$, and $\widetilde{\mathbf{n}}[k]$ being discrete r.v.s with the specified distributions.

The introduced phase shift $\varphi[k]$ is unacceptable and counter-measures actions such as a phase estimation and correction are necessary. In the following, it is assumed that this problem has been solved perfectly and we have to deal only with the attenuation due to the fading and the AWGN noise. The attenuation $\mathbf{r}$ results in a varying receive signal energy, whose statistical properties will be investigated in the following.

The instantaneous receive energy $\mathbf{E}_S^{(R_x)}[k]$ at time step k is - as a result of the the random attenuation $\mathbf{r}[k]$ - a random variable given by

$$\begin{aligned} \mathbf{E}_S^{(R_x)}[k] &= \left|\mathbf{r}[k] \cdot e^{j\cdot\varphi[k]} \cdot x[k]\right|^2 \\ &= \left|\mathbf{r}[k] \cdot e^{j\cdot\varphi[k]}\right|^2 \cdot |x[k]|^2 \\ &= \mathbf{r}^2[k] \cdot E_S^{(T_x)}[k], \end{aligned} \tag{3.11}$$

i.e. the receive and transmit energies are related by the **square** of the attenuation. For the average receive and transmit energies

$$\begin{aligned} \overline{E_S^{(Rx)}} &= E\left\{\mathbf{E}_S^{(Rx)}\right\} \\ &= \sum_{i=0}^{M-1} E\left\{\left|\mathbf{r}\cdot e^{j\cdot\varphi}\cdot\mathbf{x}\right|^2 \middle| \mathbf{x}=x_i\right\} \cdot P(\mathbf{x}=x_i) \\ &= \frac{1}{M}\cdot\sum_{i=0}^{M-1} E\left\{\mathbf{r}^2\right\}\cdot|x_i|^2 \\ &= E\left\{\mathbf{r}^2\right\}\cdot\overline{E_S^{(Tx)}} \end{aligned} \tag{3.12}$$

a similar relation holds and for transmission alphabets with constant symbol amplitudes

$$\overline{E_S^{(Rx)}} = E\left\{\mathbf{r}^2\right\}\cdot E_S^{(Tx)}.$$

Hence, the instantaneous receive energy $\mathbf{E}_S^{(R_x)}[k]$ obeys a χ^2- distribution[5] with degree of freedom 2 (see Appendix C.1.1 for details) and the average receive energy $\overline{E_S^{(Rx)}}$ is related to the average transmit energy $\overline{E_S^{(Tx)}}$ via the expected value of this distribution.

Similar to the AWGN case, if we want to avoid the distinction between the average transmit and receive energies, the distribution parameter σ in Equation 3.10 must be chosen in such a way, that $E\left\{\mathbf{r}^2\right\} = 1$. Appendix C.1.1 treats this issue and how a Rayleigh distributed r.v. $\mathbf{r}$ with this property can be generated. If this is achieved,

$$\overline{E_S^{(Rx)}} = \overline{E_S^{(Tx)}}$$

and the PDFs and the cumulative distribution functions (CDFs) of the random variables $\mathbf{E}_S^{(Rx)}[k]$ in dependency of the average receive energy (which is identical to the transmit energy) are (see Appendix C.1.1)

[5] More precisely, if $X_1, X_2, \ldots, X_n$ are n independent $N(0,1)$ distributed random variables, the random variable $X = \sum_{i=1}^{n} X_i^2$ is said to be χ^2 distributed of degree of freedom n and PDF $f_{\chi^2}^{(n)}(x)$. However, $\mathbf{r}^2$ in Equation 3.11 can be generated with two $N(0,\sigma)$ r.v.s (see Appendix C.1.1) and has therefore a stretched PDF compared to a χ^2-r.v.: $f_{\mathbf{r}^2}(x) = \frac{1}{\sigma^2} f_{\chi^2}^{(2)}\left(\frac{x}{\sigma^2}\right)$.

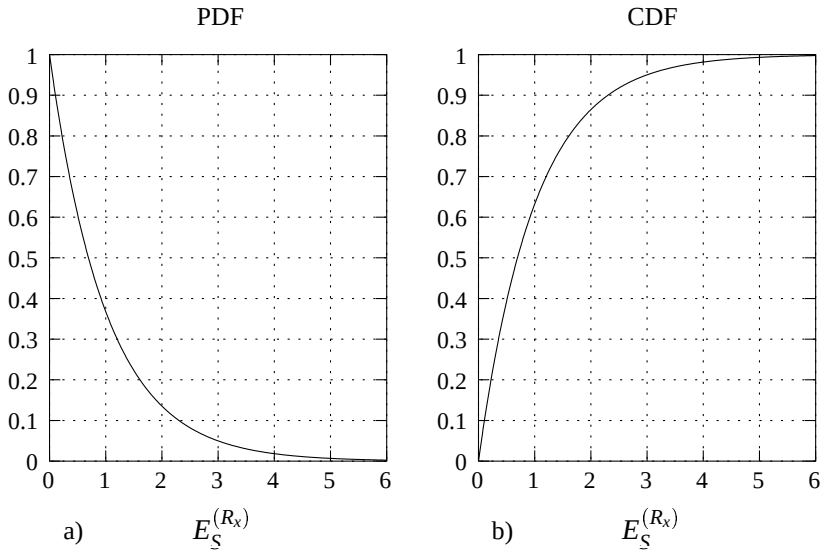

Figure 3.6: Density (a) and Distribution (b) of Received Signal Energy for $\overline{E_S^{(R_x)}} = 1$.

$$\begin{aligned} f_{\mathbf{E}_S^{(Rx)}}(x) &= \left[\overline{E_S^{(Rx)}}\right]^{-1} \cdot e^{-\left[\overline{E_S^{(Rx)}}\right]^{-1} \cdot x} \cdot u(x) \\ F_{\mathbf{E}_S^{(Rx)}}(x) &= \left(1 - e^{-\left[\overline{E_S^{(Rx)}}\right]^{-1} \cdot x}\right) \cdot u(x). \end{aligned} \tag{3.13}$$

Figure 3.6 depicts both functions over a linear $E_S^{(R_x)}$ range for $\overline{E_S^{(Rx)}} = 1$.

From 3.6 b) it can be obtained that in more than 60% of all cases the signal energy experiences an attenuation instead of an amplification. However, the destructive character of this channel becomes even more apparent if both functions are transformed into a logarithmic $E_S^{(R_x)}$ range in dB (see Appendix C.1.2 for the transformation). The PDF, shown in Figure 3.7 a), reveals that signal amplifications of 10 dB or higher hardly occur, whereas signal fades of 10 dB or lower are likely. More precisely, the CDF (Figure 3.7 b) in double log scale) reveals that in ten percent of all cases the signal collapses by more than 10 dB.

As with the received instantaneous signal energy $\mathbf{E}_S^{(R_x)}[k]$, the resulting instantaneous signal-to-noise ratio $\frac{\mathbf{E}_S^{(R_x)}}{N_0}[k]$ is a r.v. with an expected value

$$\begin{aligned} E\left\{\frac{\mathbf{E}_S^{(R_x)}}{N_0}[k]\right\} &= \frac{\overline{E_S^{(Rx)}}}{N_0} \\ &= \frac{\overline{E_S^{(Tx)}}}{N_0} \end{aligned}$$

($E\left\{\mathbf{r}^2\right\} = 1$) and a PDF and CDF of

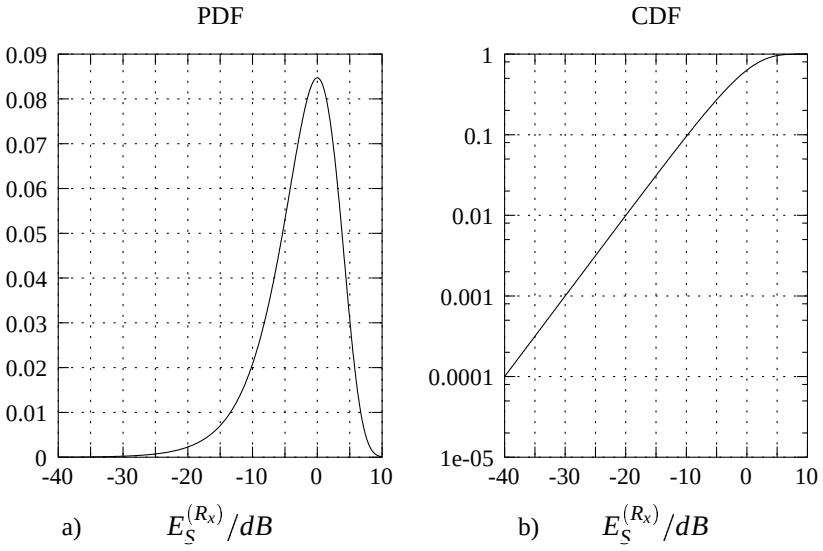

Figure 3.7: Density (a) and Distribution (b) of Received Signal Energy for $\overline{E_S^{(R_x)}} = 0\,dB$.

$$f_{\frac{\mathbf{E}_S^{(Rx)}}{N_0}}(x) = \left[\frac{\overline{E_S^{(Rx)}}}{N_0}\right]^{-1} \cdot e^{-\left[\frac{\overline{E_S^{(Rx)}}}{N_0}\right]^{-1} \cdot x} \cdot u(x)$$

$$F_{\frac{\mathbf{E}_S^{(Rx)}}{N_0}}(x) = \left(1 - e^{-\left[\frac{\overline{E_S^{(Rx)}}}{N_0}\right]^{-1} \cdot x}\right) \cdot u(x).$$

In logarithmic measure, the PDF and the CDF of $\frac{\mathbf{E}_S^{(R_x)}}{N_0}[k]$ are just shifted versions of the distributions of $\mathbf{E}_S^{(R_x)}[k]$. Hence, they show the same behavior as depicted in Figure 3.7.

3.3.2 Rayleigh Channel Model

In the AWGN section of this chapter we presented a channel model with only one degree of freedom to achieve a signal with a certain $\frac{E_S^{(Rx)}}{N_0}$ (Figure 3.3). The Rayleigh model has in its general form 3 degrees of freedom: The average transmit energy $\overline{E_S^{(T_x)}}$, the parameter σ in the Rayleigh distribution of $\mathbf{r}$, and the spectral noise power density N_0. A variation of any of these parameters, however, influences only the average receive ratio $\frac{\overline{E_S^{(Rx)}}}{N_0}$ but not its general distribution. Hence, to get rid of two redundant parameters, we agree without loss of generality on the normalized transmit symbol alphabet $\overline{x}[k] \in \{\overline{x_0}, \ldots, \overline{x_{M-1}}\}$ with average energy of one (Equation 3.7) and a Rayleigh distribution of $\mathbf{r}$ with $E\{\mathbf{r}^2\} = 1$, i.e. we set the Rayleigh parameter $\sigma = \frac{1}{\sqrt{2}}$ (compare Equation 3.10).

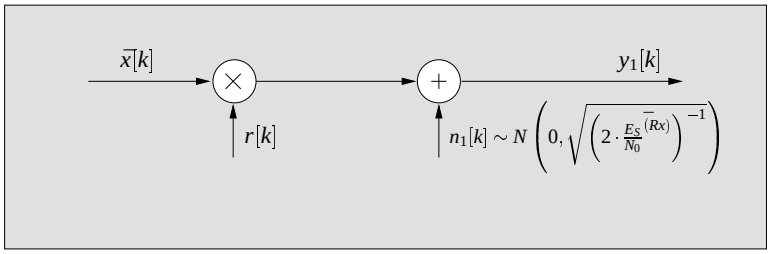

Figure 3.8: Discrete Rayleigh Model 1

Then, passing a certain symbol $\overline{x}[k]$ through the Rayleigh channel requires the execution of two random experiments:

1. The generation of a Rayleigh distributed r.v. $\mathbf{r}[k]$ with $E\left\{\mathbf{r}^2[k]\right\} = 1$, which is then multiplied with the symbol $\overline{x}[k]$ to be transmitted (modeling of the amplitude fading)
2. The generation of a complex normal distributed r.v via two independent meanfree normal distributed r.v.s with standard deviation $\sigma = \sqrt{\left(2 \cdot \frac{\overline{E_S^{(Rx)}}}{N_0}\right)^{-1}}$. The complex noise is then added to the product of 1 to form the output $\mathbf{y}[k]$.

Figure 3.8 illustrates this model.

Under the assumption that the amplitude fading $\mathbf{r} = r$ and the transmit symbol $\overline{x}$ are given, the model reduces to a simple AWGN channel with $r \cdot \overline{x}$ as deterministic part (mean of the complex distribution) and the above mentioned variance for both dimensions. Hence, the conditioned PDF of the model output $\mathbf{y}$ under these conditions is

$$f_{\mathbf{y}_1|\mathbf{r},\overline{\mathbf{x}}}(y|\, r, \overline{x}) = \frac{\overline{E_S^{(Rx)}}}{\pi \cdot N_0} \cdot e^{-\frac{\overline{E_S^{(Rx)}}}{N_0} \cdot |y - r \cdot \overline{x}|^2}. \tag{3.14}$$

The sequence of output symbols has a varying signal energy due to the Rayleigh fading, whereas the noise statistics are constant. In that sense, the Rayleigh model is nothing but an AWGN channel where the fading influences the deterministic signal part. Consequently, the amplitude fading $r[k]$ results in a varying signal-energy-to-noise-power-density ratio

$$\frac{E_S}{N_0}[k] = r^2[k] \cdot \frac{\overline{E_S^{(Rx)}}}{N_0}. \tag{3.15}$$

Therefore, the AWGN model of Figure 3.3 on Page 27 with the varying ratio of Equation 3.15 is also an equivalent Rayleigh model.

This model, depicted in Figure 3.9, creates samples $\mathbf{y}[k]$ with an unchanged deterministic part and a varying SNR. Again, two consecutive random experiments are necessary:

1. An instant $r[k]$ of Rayleigh distributed r.v. $\mathbf{r}$ is generated, squared and multiplied with the intended $\frac{\overline{E_S^{(Rx)}}}{N_0}$ resulting in a certain signal-power-to-noise-power-density ratio $\frac{E_S}{N_0}[k]$.

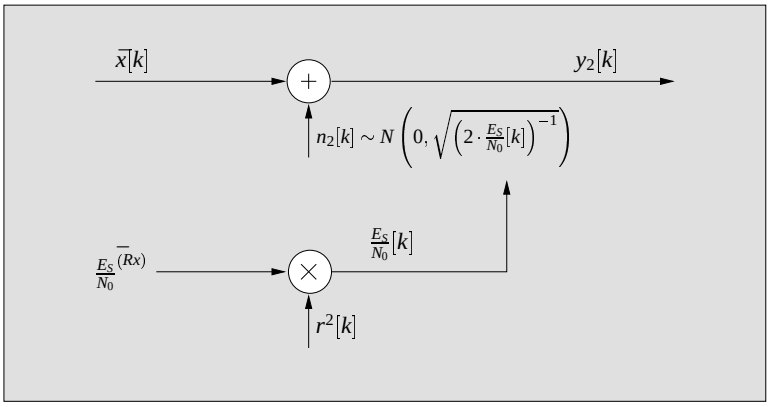

Figure 3.9: Discrete Rayleigh Model 2

2. Dependent on the outcome of step 1, a mean free normal distributed r.v. with a variance equal to $\left[2 \cdot \frac{E_S}{N_0}[k]\right]^{-\frac{1}{2}}$ is generated and added to the symbol $\overline{x}[k]$ to be transmitted, yielding the receive sample $\mathbf{y}[k]$.

As a result of the different generation, the PDF of the second Rayleigh model output has an altered PDF for a given fading r and transmit symbol $\overline{x}$ compared to model 1. To be more precise, the mean of the complex distribution is $\overline{x}$ and the squared variance equal to

$$\sigma^2 = \left[2 \cdot \frac{E_S}{N_0}[k]\right]^{-1} = \left[2 \cdot r^2[k] \cdot \frac{\overline{E_S^{(Rx)}}}{N_0}\right]^{-1}$$

and for the conditioned PDF follows

$$f_{\mathbf{y}_2|\mathbf{r},\overline{\mathbf{x}}}(y|\, r, \overline{x}) = \frac{r^2}{\pi} \cdot \frac{\overline{E_S^{(Rx)}}}{N_0} \cdot e^{-r^2 \cdot \frac{\overline{E_S^{(Rx)}}}{N_0} \cdot |y - \overline{x}|^2}. \tag{3.16}$$

Although the output of both models show Rayleigh characteristics and have an average signal-energy-to-noise-power-density ratio of $\frac{\overline{E_S^{(Rx)}}}{N_0}$, they require a different treatment for maximum likelihood decoding and for combining. These issues will be treated in Chapter 5.1.

3.3.3 Temporal Correlation

So far, only the **distribution** of the instantaneous signal-energy-to-noise-power-density ratio has been considered. Above, the distribution of the receive energy $\mathbf{E}_S^{(Rx)}$ and the corresponding ratio $\frac{\mathbf{E}_S^{(Rx)}}{N_0}$ of a Rayleigh channel was given. That is, we expect the receive signal energy to obey this distribution if measurements are made in **arbitrary** locations. If the environment remains unchanged, i.e. neither transmitter, nor receiver, nor the obstructions responsible for the channel characteristics move, and also the transmit frequency remains unchanged, we have a constant

energy amplification due to the channel. However, any slightest movement of the mentioned objects results in a possibly dramatic change of the receive energy. Clearly, since any movements are continuous, also the change of $\mathbf{E}_S^{(Rx)}$ will be a continuous process. Due to the high degree of freedoms, this process is extremely difficult to model. The usual way utilizes a statistical model such as the wide sense stationary uncorrelated scattering (WSSUS) model. In Section 3.4 such a statistical model is discussed. In the remainder of this section, however, only two simple special cases are considered, namely the so-called Constant Rayleigh and Independent Rayleigh channel.

Constant Rayleigh Channel

For packetized transmission schemes a simple way to model a varying fading is to assume a constant fading for the durations of the packets and uncorrelated fading between the individual packets. Although these assumptions do not seem to lead to a useful model, they nicely fit into the context of ARQ systems: Many performance parameters, which will be defined and investigated in Chapter 4, are only dependent on the correlation of the original packet and its possible retransmissions, whereas the correlation of succeeding packets are of no concern. Hence, this channel model channel represents the marginal case of an extreme slow motion of or within the environment (slow enough so that the channel does not change noticeably over a single transmission) and a very long time until the retransmission request in an ARQ system arrives (long enough so that the channel is uncorrelated, despite the slow motion).

> The constant Rayleigh channel assumes a constant receive energy $\mathbf{E}_S^{(Rx)}[k]$ for a certain transmission and statistically mutually uncorrelated channels for all pertaining retransmissions. The amplitude fading of all transmissions are Rayleigh distributed.

The author likes to emphasis, that this model does not make any specific statements about the correlation of consecutively received transmissions. Obviously, due to the slow motion, these transmissions are highly correlated. Some of the results, which will be obtained in Chapter 4, are independent of these intertransmission correlations and the corresponding results are applicable to this channel. On the other hand, some derivations assume succeeding packets to be uncorrelated and results obtained with this channel model are little meaningful.

The constant Rayleigh channel is in fact the most destructive channel presented in this work. The destructive nature can be illustrated with a small example and Figure 3.7: Suppose a packet based FEC system is designed to work down to 10 dB over an AWGN channel with an acceptable performance and it is used without any modifications over a constant Rayleigh channel. Then, only if the **receive** signal-energy-to-noise-power-density ratio is above the crucial value of 10 dB an acceptable performance is expected and for values below that the systems fails. With the constraint of being operational in 99 % per cent of the times, we can obtain from Figure 3.7 that the lowest admissible $\overline{\frac{E_S^{(Rx)}}{N_0}}$ would be 30 dB. More drastically, if the system may fail only at a rate of 10^{-6}, the lowest possible $\overline{\frac{E_S^{(Rx)}}{N_0}}$ is 70 dB. As a consequence, without any counteracting the system needs to be extremely overdesigned to ensure a certain probability of operativeness for a constant Rayleigh channel. ARQ systems are one way to overcome these effects.

Independent Rayleigh Channel

As the speed of the motion of or within the environment increases, the time it takes until first transmission and pertaining retransmission can be regarded as independent decreases. At a certain speed even the channel statistics over a single transmission might change even to that point where some symbols in a transmission are independent of others. In a practical system, this effect is emphasized with the use of a symbol interleaver. In fact, if the interleaver length is long enough all symbols of a specific transmission can be regarded as independent and consequently also the correlation between succeeding packets is zero.

The described scenario is modeled by the independent Rayleigh channel: The receive energies $\mathbf{E}_S^{(Rx)}[k]$ of all transmitted **symbols** are mutual independently Rayleigh distributed. Hence, all transmissions are also mutually independent corrupted and all results of Chapter 4 can be applied to this channel model.

The performance of digital communications system over this channel is better than over the constant Rayleigh channel, especially if error correction codes are employed.

The Rayleigh distribution was obtained under the assumption that we do have a single non-direct path. The following section considers a more general scenario (see Figure 3.4), where several non-direct main paths between transmitter and receiver exist.

3.4 Multiple Rayleigh Channels

In the multipath propagation environment, as depicted in Figure 3.4 on Page 28, the receiver is provided with a certain number l of signal replicas. Also, with exception of the direct path, the main paths can be modeled as independent Rayleigh channels, so that the receiver is supplied with l independent Rayleigh corrupted copies of the transmit signal. Unfortunately, the copies arrive with a temporal delay and, hence, a mixture of these replicas is received. Yet, if the signal bandwidth is much greater than the coherence bandwidth of the channel, the multipath components can be resolved (see [Pro95] for details). As early as 1958, an optimum receiver for this environment was advised by Price and Green [PrGr58], namely the RAKE receiver.

Besides the multipath environment, there are may ways in which further copies can be supplied. Examples include the transmission of the signal over l carriers (*frequency diversity*) as in OFDM (*Orthogonal Frequency Division Multiplexing*), the usage of multiple antenna (*antenna diversity*), the usage of the polarization (*polarization diversity*), and the time delayed retransmission (*time diversity*). Yet, there is a difference between the diversity due to multipath and the other mentioned diversity schemes: The multipath environment solely provide l **replicas**, whereas in the other diversity schemes one has the freedom to transmit **any redundant information** to aid the decoding process.

3.4.1 Maximum Ratio Combining

As one result of this work, repetitions as retransmissions are not an optimal way to provide further information. If one still decides to do so, or if only copies can be sent (as it is the case in an multipath environment) an optimum way to combine the repetitions has already been published by Brennan [Bre59], namely maximum-ratio-combining (MRC). This principle of diversity combining is discussed in detail in Section 5.1.4. For now, we will only use the results to obtain an equivalent channel model for the multipath environment:

> A maximum-ratio-diversity combining scheme combines l independent **replicas** of a signals received with the l individual signal-energy-to-noise-power-density ratios $\left.\frac{E_S^{(Rx)}}{N_0}\right|_j$, $j = 0, \ldots, l-1$ to the same signal with an effective signal-energy-to-noise-power-density ratio of

$$\left.\frac{E_S^{(Rx)}}{N_0}\right|_{MRC} = \sum_{j=0}^{l-1} \left.\frac{E_S^{(Rx)}}{N_0}\right|_j \tag{3.17}$$

As an example, if the channel consists of two paths, which have at a certain time instant a signal-energy-to-noise-power-density ratios of $0\,dB$ and $3\,dB$, respectively, the maximum ratio combined signal has a signal-energy-to-noise-power-density ratio of

$$\begin{aligned} \left.\frac{E_S^{(Rx)}}{N_0}\right|_{MRC} &= 10 \cdot log\,(1+2)\ dB \\ &= 4.87\,dB. \end{aligned}$$

If the individual paths of a multipath environment observe amplitude fading $\mathbf{r}_j[k]$, $j = 0, \ldots, l-1$, which is slow enough to perform MRC, it follows for the the average receive ratio

$$\begin{aligned} \left.\overline{\frac{E_S^{(Rx)}}{N_0}}[k]\right|_{MRC} &= E\left\{\sum_{j=0}^{l-1} \left.\frac{E_S^{(Rx)}}{N_0}[k]\right|_j\right\} \\ &= \sum_{j=0}^{l-1} E\left\{\left.\frac{E_S^{(Rx)}}{N_0}[k]\right|_j\right\} \\ &= \sum_{j=0}^{l-1} \left.\overline{\frac{E_S^{(Rx)}}{N_0}}[k]\right|_j \end{aligned} \tag{3.18}$$

Hence, MRC leads to an increased in the average receive signal-energy-to-noise-power-density ratio. Yet, the performance of a digital communication system transmitting over a multiple Rayleigh channels not only depends on the average receive signal-energy-to-noise-power-density ratio, but also on the ratios of the individual signal-energy-to-noise-power-density ratios $\left.\overline{\frac{E_S^{(Rx)}}{N_0}}[k]\right|_j$.

For its characterization we introduce the ratios

$$p_j = \left.\overline{\frac{E_S^{(Rx)}}{N_0}}\right|_j \Big/ \left.\overline{\frac{E_S}{N_0}}\right|_{Norm},$$

where $\left.\overline{\frac{E_S}{N_0}}\right|_{Norm}$ represents a normalization ratio. Table 3.1 on Page 42 lists these ratios in dB scale with

$$\left.\overline{\frac{E_S}{N_0}}\right|_{Norm} = max_{j=0,\ldots,l-1}\left\{\left.\overline{\frac{E_S^{(Rx)}}{N_0}}\right|_j\right\}$$

for the UMTS standardization channel models.

The instant receive ratio $\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k]$ of a multiple Rayleigh channel with the ratio profile $\{p_0, p_1, \ldots, p_{l-1}\}$ is given by

$$\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k] = \sum_{j=0}^{l-1} p_j \cdot \mathbf{r}_j^2[k] \cdot \overline{\frac{E_S^{(Tx)}}{N_0}}.$$

where the $\mathbf{r}_j$ are mutually independent Rayleigh processes. Then, with $E\left\{\mathbf{r}_j^2[k]\right\} = 1$, the instantaneous receive ratio $\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k]$ has an expected value of

$$\begin{aligned}\overline{\frac{E_S^{(Rx)}}{N_0}}[k] &= E\left\{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k]\right\} \\ &= \overline{\frac{E_S^{(Tx)}}{N_0}} \cdot \textstyle\sum_{j=0}^{l-1} p_j\,.\end{aligned}$$

Again, if one wants to avoid the distinction between average transmit and receive ratios, normalized ratio coefficients

$$\begin{aligned}\overline{p_k} &= \frac{p_k}{\sum_{j=0}^{l-1} p_j} \\ &= \frac{\left.\overline{E_S^{(Rx)}}\Big/N_0\right|_k}{\sum_{j=0}^{l-1}\left.\overline{E_S^{(Rx)}}\Big/N_0\right|_j}\end{aligned}$$

are introduced and the random instantaneous signal-energy-to-noise-power-density ratio obtained via

$$\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k] = \sum_{j=0}^{l-1} \overline{p_j} \cdot \mathbf{r}^2[k] \cdot \overline{\frac{E_S^{(Tx)}}{N_0}}$$

has the statistical characteristics of a channel with the ratio profile $\{p_0, \ldots, p_{l-1}\}$ and an expected value of

$$E\left\{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k]\right\} = \overline{\frac{E_S^{(Tx)}}{N_0}}.$$

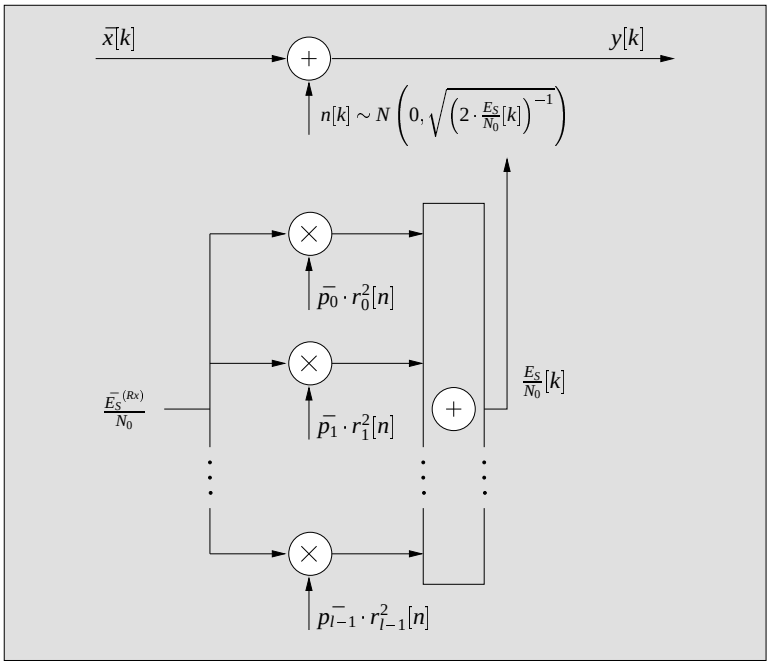

Figure 3.10: Discrete Channel Model of a WSSUS Channel with Optimum Channel Estimation and Combining.

Hence, using l mutually independent Rayleigh processes and the normalized ratio profile a channel model for the multiple Rayleigh channel with ideal MRC can be constructed in analogy to the discrete Rayleigh model depicted in Figure 3.9 (see Figure 3.10).

Consequently, the output has a variance of

$$\sigma^2 = \left[2 \cdot \frac{E_S}{N_0}[k]\right]^{-1} = \left[2 \cdot \sum_{j=0}^{l-1} \overline{p_j} \cdot \mathbf{r}_j^2[k] \cdot \overline{\frac{E_S^{(Tx)}}{N_0}}\right]^{-1}$$

and the PDF that the channel output $\mathbf{y} = y$ is obtained under the condition that $\overline{x} = \overline{x}$ was transmitted and the channel state information $r_0, \ldots, r_{l-1}$ are known is

$$f_{\mathbf{y}|\overline{x},\mathbf{r}_0,\ldots,\mathbf{r}_{l-1}}(y|\,\overline{x}, r_0, \ldots, r_{l-1}) = \frac{1}{\pi} \cdot \sum_{j=0}^{l-1} \overline{p_j} \cdot \mathbf{r}_j^2[k] \cdot \overline{\frac{E_S^{(Tx)}}{N_0}} \cdot e^{-\sum_{j=0}^{l-1} \overline{p_j} \cdot \mathbf{r}_j^2[k] \cdot \overline{\frac{E_S^{(Tx)}}{N_0}} \cdot |y-\overline{x}|^2}, \quad (3.19)$$

or using the resulting signal-energy-to-noise-power-density $\frac{E_S}{N_0}$

$$f_{\mathbf{y}|\overline{x},\frac{\mathbf{E_S}}{\mathbf{N_0}}}\left(y|\,\overline{x}, \frac{E_S}{N_0}\right) = \frac{1}{\pi} \cdot \frac{E_S}{N_0} \cdot e^{-\frac{E_S}{N_0} \cdot |y-\overline{x}|^2}. \quad (3.20)$$

The normalized energy ratio profile $\{\overline{p_0}, \ldots, \overline{p_{l-1}}\}$ of the l main paths heavily depends on the diversity scheme and the transmission conditions. As a special case, all energy ratios may be identical

$$\overline{p_j} = \frac{1}{l},\ j = 0, \ldots, l-1\,,$$

while in the general case they have different values. As an example, both copies in a system with polarization diversity are usually received with the same energy ratio $\frac{1}{2}$, whereas OFDM transmitted over a frequency selective channel may result in different average attenuations for the l replicas.

In order to study the behavior of a digital communication system used over a multiple Rayleigh channel, we start with the special case of l channels with normalized energy ratios $\frac{1}{l}$ before we move on the the more general case

3.4.2 Constant Normalized Energy Ratio Profile

A constant normalized energy ratio profile $\overline{p_j} = \frac{1}{l},\ j = 0, \ldots, l-1$ is the result of l statistically equivalent channels. Such a system is known as a Lth-order diversity system (with $L = l$; the variable L is in this work reserved for other use, hence it will be called lth-order diversity scheme) [Pro95]. Then, MRC is a simple unscaled addition of all received replicas and the r.v. $\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k]$ has a scaled χ^2 distribution with degree of freedom $2 \cdot l$. In Appendix C.2.1 it is shown that in this case the corresponding PDF is

$$f_{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}}(x) = \frac{1}{\alpha^l \cdot \Gamma(l)} \cdot x^{l-1} \cdot e^{-\frac{x}{\alpha}} \cdot u(x)\,,\ l \geq 1$$

and the CDF

$$\begin{aligned} F_{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}}(x) &= \Gamma_{inc}\left(\tfrac{x}{\alpha}, l\right),\ x \geq 0,\ l \geq 1 \\ &= \left[1 - e^{-\frac{x}{\alpha}} \cdot \textstyle\sum_{k=0}^{l-1} \frac{1}{k!}\left(\frac{x}{\alpha}\right)^k\right] \cdot u(x)\,,\ l = 1, 2, \ldots \end{aligned}$$

with

$$\alpha = \frac{1}{l} \cdot \frac{\overline{E_S^{(Tx)}}}{N_0}$$

and $\Gamma_{inc}(x, n)$ denoting the incomplete Gamma function[6] [Abr65]

$$\Gamma_{inc}(x, n) = \frac{1}{\Gamma(n)} \cdot \int_0^x v^{n-1} \cdot e^{-v} dv.$$

The various CDFs for different diversity levels l are depicted over linear and double-log scales in Figure 3.11. The plots reveal the dramatic performance gains with increasing diversity l com-

[6] The incomplete Gamma function can be obtained via numerical tables or more easily and precisely with MATLAB via the function *gammainc(x,n)*.

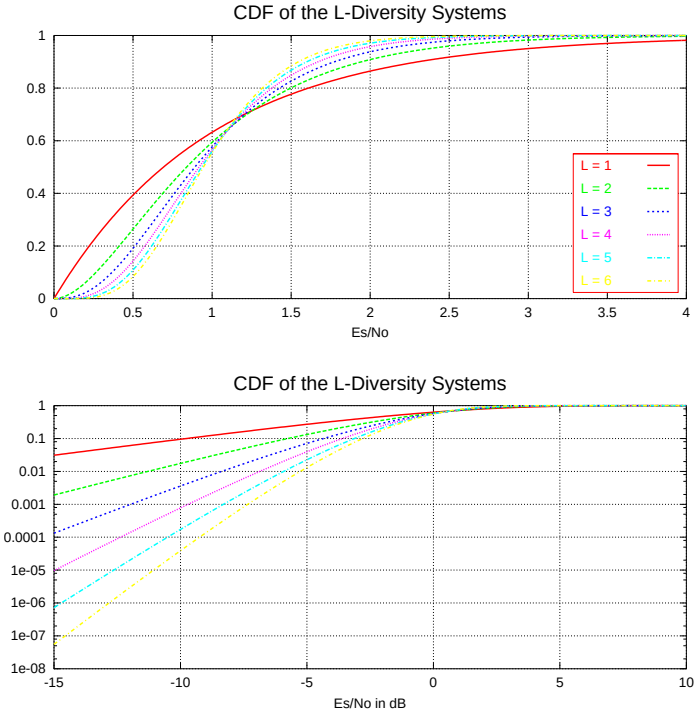

Figure 3.11: CDF of the L-th Order Diversity Schemes for $\frac{\overline{E_S^{(Tx)}}}{N_0} = 1$.

pared to the Rayleigh channel ($l = 1$). As an example in a system with diversity level $l = 4$, the receive energy drops 10 dB below the average receive energy in less than 10^{-3} times. Over a Rayleigh channel, however, the receive energy is in 1 out of 10 times 10 dB below the average receive ratio.

3.4.3 General Energy Ratio Profile

Now we want to assume a normalized energy ratio profile $\{\overline{p_0}, \ldots, \overline{p_{l-1}}\}$ in which all ratios are mutual different. As shown in Appendix C.2.2, the corresponding PDF and CDF are

$$f_{\mathbf{E}_S^{(Rx)}}(x) = \sum_{i=1}^{n} \frac{\alpha_i^{n-2} \cdot e^{-\frac{x}{\alpha_i}}}{\prod_{j=1, j\neq i}^{n} (\alpha_i - \alpha_j)} \cdot u(x), \; n \geq 2$$

$$F_{\mathbf{E}_S^{(Rx)}}(x) = \sum_{i=1}^{n} \frac{\alpha_i^{n-1} \left(1 - e^{-\frac{x}{\alpha_i}}\right)}{\prod_{j=1, j\neq i}^{n} (\alpha_i - \alpha_j)} \cdot u(x), \; n \geq 2$$

with

$$\alpha_i = \overline{p_i} \cdot \frac{\overline{E_S^{(Tx)}}}{N_0}.$$

Tap	Indoor / Office A	Indoor / Office B	Outdoor To Indoor A
1	0.0	0.0	0.0
2	-3.0	-3.6	-9.7
3	-10.0	-7.2	-19.2
4	-18.0	-10.8	-22.8
5	-26.0	-18.0	-
6	-32.0	-25.2	-
Tap	Outdoor To Indoor B	Vehicular A	Vehicular B
1	0.0	0.0	-2.5
2	-0.9	-1.0	0.0
3	-4.9	-9.0	-12.8
4	-8.0	-10.0	-10.0
5	-7.8	-15.0	-25.2
6	-23.9	-20.0	-16.0

Table 3.1: Energy Profiles in dB of the Test Environments, Specified for UMTS.

Hence, the PDF and the CDF heavily depend on the normalized energy ratio profile $\{\overline{p_0}, \ldots, \overline{p_{l-1}}\}$. The closer they are to an equal distribution, the closer we get to the l - diversity case, which achieves the best performance among all energy ratio profiles $\{\overline{p_0}, \ldots, \overline{p_{l-1}}\}$. This can be argued as follows: If $l-1$ normalized energy ratios are approximately equal $\overline{p_j} \simeq \frac{1}{l-1}$, $j = 0, \ldots, l-2$ and the remaining one is much smaller than the others, i.e. $\overline{p_{l-1}} \simeq 0$, then the performance if identical to that of a $(l-1)$-diversity scheme.

Table 3.1 depicts the (unnormalized) energy profiles for the test environments specified for the UMTS standardization [UMTS98]. From the above said, we expect a digital system operating over the Outdoor To Indoor B environment to have the best performance among the given UMTS channels, since its energy profile distribution is closest to the uniform distribution. The Outdoor to Indoor A channel, on the other hand, will show the worst performance, since firstly it has only 4 taps, but also since the first main path heavily dominates over all other paths.

Figures 3.12 and 3.13 depict the PDF and CDF of the various UMTS channels of Table 3.1 in linear and double log-scale. As expected, Outdoor To Indoor B channel has the narrowest PDF while the Outdoor To Indoor B channel has the widest, yielding to the worst performance. Figure 3.13 reveals how different the CDF of channels with the same number of taps can be. In fact, the performance of the Outdoor to Indoor A channel is only marginally better than that of a Rayleigh channel (compare Figures 3.7 on Page 32 with 3.13).

Up to now, we treated the PDF and the CDF of instants of the receive energy $\mathbf{E}_S^{(Rx)}[k]$ or the receive signal-power-to-noise-power-density ratio $\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k]$ of l Rayleigh channels with optimum combining and arbitrary distribution of their energy profile. We also showed how a random variable with such a distribution can be generated with the help of L independent instants of Rayleigh distributed random variables $\mathbf{r}_j[k]$, $j = 0, \ldots, l-1$. Yet we did not treat the issue of the

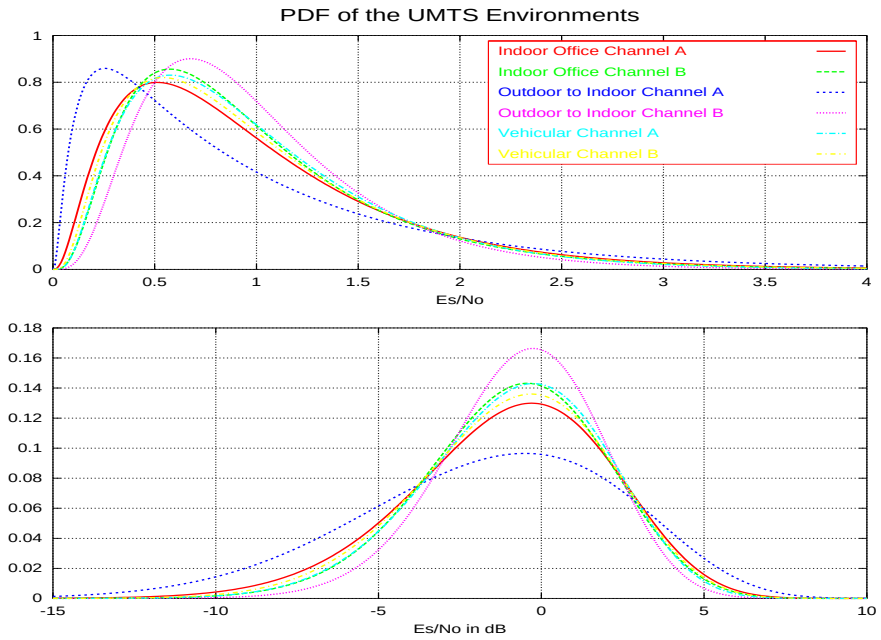

Figure 3.12: PDF of Various Mobile Channels specified for UMTS for $\frac{\overline{E_S^{(Tx)}}}{N_0} = 1$.

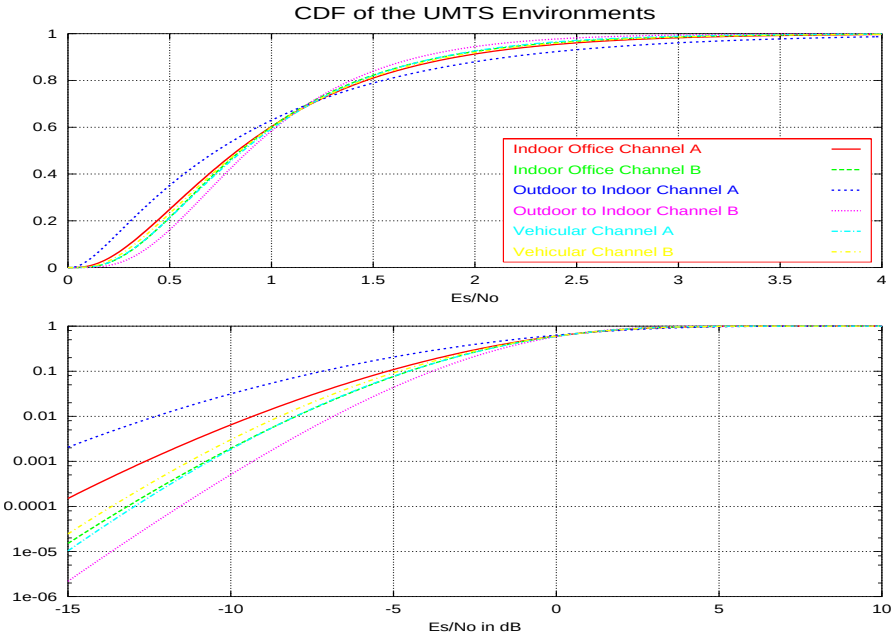

Figure 3.13: CDF of Various Mobile Channels specified for UMTS for $\frac{\overline{E_S^{(Tx)}}}{N_0} = 1$.

temporal correlation between succeeding instants of the r.v.s $\mathbf{E}_S^{(Rx)}[k]$ and $\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}[k]$.

3.4.4 Temporal Correlation

The topic of the temporal correlation of Rayleigh random variables was treated only briefly in Section 3.3. We covered the special cases of constant correlation within a single transmission and zero correlation between different transmissions (Constant Rayleigh channel) and the case of zero correlation between individual symbols and hence also between packets (Independent Rayleigh). In this subsection we want to look into a continuous correlation and show under which circumstance the assumptions of a packetwise constant correlation or a symbolwise zero correlation is justified.

In order to do so, we need a statistical model of the temporal correlation. In UMTS, the well known wide sense stationary uncorrelated scattering (WSSUS) model is specified as test environment for development . As mentioned, in this model, the signal consists of L independent Rayleigh channels with a specified normalized energy profile $\{\overline{p_0}, \ldots, \overline{p_{l-1}}\}$ and a certain delay profile, which is of no concern for us since we assume an optimum receiver (RAKE receiver).

The modeling of the correlation of time instants of a specific main path (different main paths are uncorrelated at any time) is achieved by modeling the corresponding attenuation sequence $\mathbf{r}(t) \cdot e^{-j\cdot\varphi(t)}$ (see Figure 3.5) as time variant process. The power spectrum of this complex time invariant process, called the Doppler spectrum, controls the amount of correlation and hence the rate of fading. The actual Doppler spectrum depends on the velocity of the mobile, the number of receive rays, and their geometrical relation of the mobile. For the following assumptions, however, specific Doppler spectra can be obtained:

1. A very large number of receive rays arrive uniformly distributed in azimuth at the mobile station at zero elevation.

2. A very large number of receive rays arrive uniformly distributed in elevation and azimuth at the mobile station.

Assumption 1 results in a so-called Jakes spectrum given by

$$P(f) = \frac{1}{\pi} \cdot \frac{1}{\sqrt{\left(\frac{v}{\lambda}\right)^2 - f^2}}, |f| < \frac{v}{\lambda}, \tag{3.21}$$

where v is the velocity of the mobile and λ is the wavelength of the carrier.

Assumption 2 results in a spectrum that is near flat and usually the assumption of a flat spectrum

$$P(f) = \frac{\lambda}{2 \cdot v}, |f| < \frac{v}{\lambda} \tag{3.22}$$

is made to represent this environment (see for example [UMTS98]). Figure 3.14 depicts both spectra specified for the UMTS test environments.

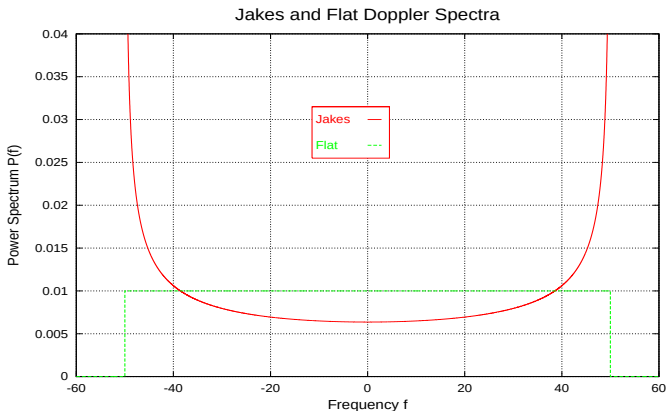

Figure 3.14: Jakes and Flat Doppler Spectra for a Mobile Velocity of 27 $\frac{km}{h}$ and a Carrier Frequency of 2 GHz (maximum Doppler frequency 50 Hz).

UMTS Test Environment	Doppler Spectrum
Indoor / Office	Flat
Outdoor To Indoor	Jakes
Vehicular	Jakes

Table 3.2: Doppler Spectra Type of the Taps of the UMTS Test Environments.

A time variant process with a certain power spectrum can be generated by passing a white process through a filter with the appropriate impulse response. For simulations in this work, however, the discrete model outlined in [Hoe90] is used. There, the received signal is modeled by the superposition of a large number of independent rays with uniform distributed phase and an Doppler frequency distributed according to the specified Doppler spectrum.

For the UMTS test environments all taps of a specific environment have identical Doppler spectra - either Jakes or Flat. Table 3.2 lists the corresponding Doppler spectrum types. For the outdoor environments Outdoor To Indoor A & B and Vehicular A & B, the Jakes spectrum was chosen, since in these environment incoming rays are received with an elevation angle close to zero. For the indoor environment Indoor / Office the Flat spectrum was chosen since the incoming rays are also reflected on the floors and the ceilings of the rooms and a uniform distribution of the elevation angle of the incoming rays is more likely than a constant zero angle.

After a decision has been made concerning the Doppler spectrum and the transmit carrier frequency, the model has only a single parameter, namely the mobile velocity v. As Equations 3.21 and 3.22 suggest, the higher the mobile velocity, the higher the maximum Doppler frequency and hence, the lower the correlation of preceding fading samples. Also from Figure 3.14 it can be seen that although the Jakes and the Flat spectrum have the identical maximum Doppler frequency, the noise process with Jakes spectrum has a lower correlation.

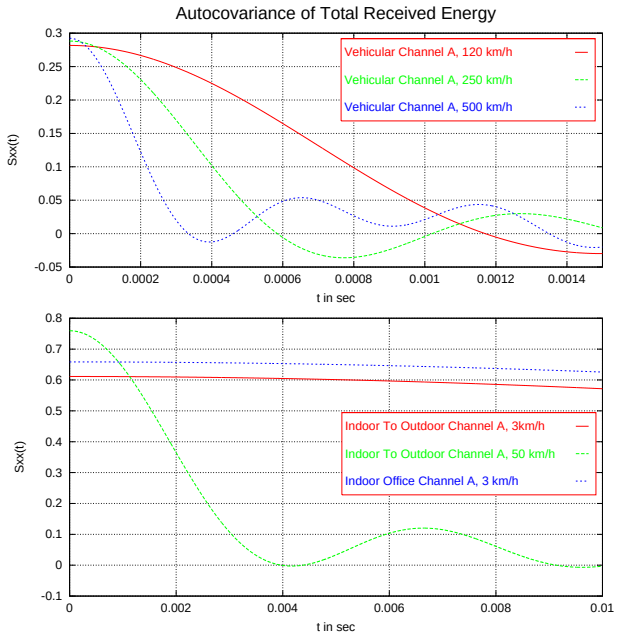

Figure 3.15: Autocovariance of Some UMTS Test Environments at Specified Mobile Velocities.

So far the correlation of noise process samples of a specific main path was treated. If an optimum receiver for this environment with optimal channel estimation is employed, the overall receive signal energy is a random process composed of the weighted sum of l independent squared random processes. The issue of the correlation of these energies $\mathbf{E}_S\,[k]$ (and hence $\frac{\mathbf{E}_S}{\mathbf{N}_0}\,[k]$ samples) was treated in this work by simulations only, rather by a theoretical analysis.

Figure 3.15 depicts simulations of the autocovariance for the various UMTS environments with the specified mobile velocities.

For the high speed downlink packet access (HSDPA) mode of UMTS [UMTS98] a PDU duration of 2 $msec$ is specified[7]. It can be seen that for the pedestrian velocity of 3 km/h the autocovariance of the Outdoor To Indoor and the Indoor / Office channel remains constant. Hence, for these environments, the assumption of a constant fading over the PDU is justified. For the vehicular channels with velocities 120 km/h, 250 km/h, and 500 km/h, the fading at the beginning of the PDU is totally uncorrelated to the fading at the end. Hence, with a symbol interleaver, the assumption of symbolwise independent fading is justified. Only the Outdoor To Indoor environment with the specified velocity of 50 km/h, the autocovariance is neither constant, nor zero for the duration of a PDU. Yet if the correlation is taken to be constant over the PDU length, the results obtained with this assumption are slightly worse and the actual performance is somewhat better.

[7] The standardization of HSDPA was by the time of writting in early 2002 still in process and the mentioned time of 2 $msec$ results from a frame duration of 10 $msec$ with 15 slots and 3 consecutive slots per HSDPA PDU.

Chapter 4

Performance Measures for ARQ Systems

In this chapter various performance measures for ARQ systems are defined. Among them are the maximum and the average number of transmissions, the throughput, various delays, memory requirements, and finally the average data rate. All these performance measures will be bounded with the help of so-called rejection probabilities, defined in Section 4.2, and other system parameters such as packet lengths, round trip delays, etc. Also, the interaction between the various performance measues and the system parameters are derived.

4.1 Coding-/ Modulation Rate and $\frac{E_S}{N_0}$ vs. $\frac{E_b}{N_0}$

Before we discuss adequate performance measures for ARQ systems, we have to answer the question versus which parameter they should be plotted. Of course, any performance heavily depends on the used channel model (see Section 3.2) and its condition. Hence, a performance plot is only valid for a specific channel model and should show the variation of that particular performance measure versus the channel condition. The apparent parameter describing the condition of the channel is the (average) $\frac{E_S}{N_0}$ ratio. In Section 3.2 this ratio was defined as the (average) quotient of the energy of the *received signal* E_S to the noise power density N_0. In that definition, the actual shape of the waveform and especially the size of the modulation alphabet is of no concern. However, if coding and/or multilevel modulation are employed, the average amount of information bits contained in each transmitted signal is dependent on the particular used scheme.

> Any performance measures compared versus $\frac{E_S}{N_0}$ assumes constant conditions on a **symbol** basis. That means the compared systems have the same symbol duration and hence approximately the same bandwidth requirements but possibly different information rates.

In almost all digital communication systems some form of error correction schemes are present. In Section 1.2, FEC systems were introduced in which redundancy in form of additional bits is added to the information bit stream. As a measure of the amount of redundancy introduced by the code, the code rate R_C is defined:

> The code rate R_C of a FEC system is defined as the number of information bits b_{info} per encoding step divided by the number of total bits b_{total} generated by the encoder:
>
> $$R_C = \frac{b_{info}}{b_{total}}$$

In schemes were the encoding procedure is done on a block basis, R_C is simply the length of the uncoded information block L_{info} divided by the codeword length L (if both words are over the same alphabet), whereas in convolutional encoded schemes, the code rate is the number of bits entering the shift registers divided by the number of output bits. Clearly,

$$R_C \leq 1.$$

However, additional bits are not the only way of introducing redundancy. Ungerboeck [Ung82] showed how it can be incorporated in the modulation sequence by using a rate $R_C = \frac{k}{k+1}$ convolutional code and an increased modulation alphabet of size $M = 2^{k+1}$. By that, the additional introduced bits are "hidden" in the increased modulation set.

In general, FEC can be used in conjunction with a multilevel modulation. Depending on the code rate R_C and the modulation alphabet size M, a modulation symbol carries a certain amount of information bits. As a measure for the average amount of information bits per transmitted symbol we introduce the coding/modulation rate R_{CM} of a FEC system:

$$R_{CM}^{FEC} = R_C \cdot log_2\left(M\right).$$

For example, an uncoded binary transmission system has a coding-/modulation rate of 1($R_C = 1$ and $M = 2$). This coding-/modulation rate is also achieved by a rate $\frac{1}{2}$ coded QPSK transmission system ($R_C = 1/2$ and $M = 4$).

As an example for a $\frac{E_S}{N_0}$-plot consider the BER curves of two convolutional encoded FEC systems (see Figure 4.1), one with code rate $R_c = \frac{1}{2}$, constraint length[1] $v = 2$ and one with $R_c = \frac{1}{4}$, $v = 8$, both using BPSK over an AWGN channel. It can be seen that the $R_c = \frac{1}{4}$ code outperforms the other code over the complete $\frac{E_S}{N_0}$-range in terms of bit error performance. However, it is not immediately apparent that this code also has only half the information rate compared to the $R_C = \frac{1}{2}$ code. Hence, the corresponding code rate as well as the level of the used modulation format is of vital information for a correct interpretation of such a plot and should therefore always be mentioned.

[1] The constraint length v is used ambiguously throughout literature. We follow the definition in [Bos98] and define the constraint length as total number of memory elements in the encoder. Hence, to corresponding trellis has 2^v states.

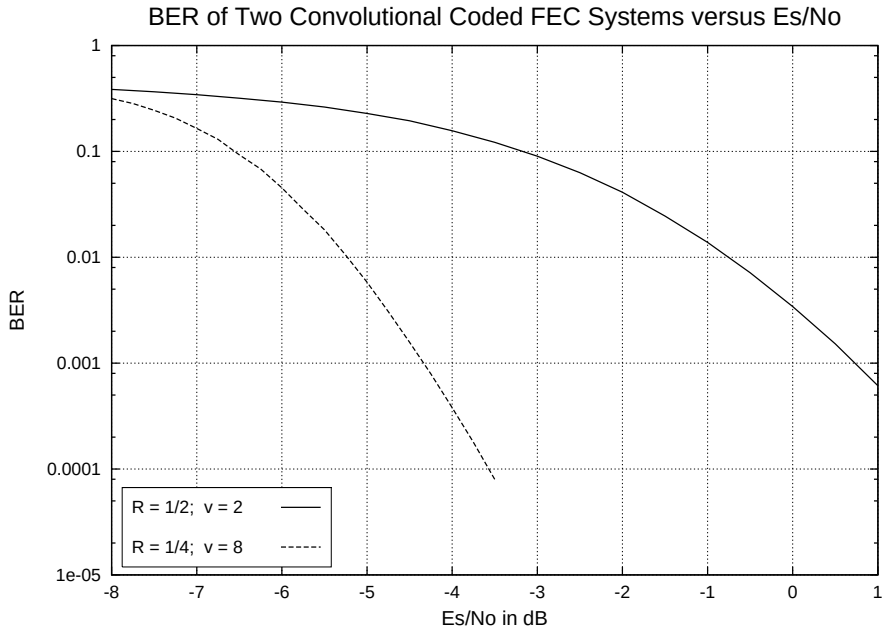

Figure 4.1: BER of Two Softdecision Decoded Convolutional Encoded FEC Systems Using BPSK over AWGN.

For block based FEC systems, the coding-/modulation rate can be immediately derived from the information length L_{info} in bit and the length of the transmitted word L in $symbols$:

$$R_{CM}^{FEC} = \frac{L_{info}}{L}. \tag{4.1}$$

In analogy with FEC systems, also for ARQ systems coding-/modulation rates can be defined. The ARQ system is a block based transmission system, which tries to transmit the information block of length $L_{info}\, bit$ via - if necessary - several transmissions of packets with possibly varying packet lengths $L_j\ symbol,\ j \geq 1$ over a channel. Therefore, we carefully need to distinguish between the coding-/modulation rates of the j-th transmission

$$R_{CM}^{FEC,j} = \frac{L_{info}}{L_j}$$

and the overall coding-/modulation rates $R_{CM}^{ARQ,j}$ of an ARQ system after the j-th transmission. The coding-/modulation rate of the ARQ system after the j-th transmission takes all sent packets up to the present into account. Hence, it is defined as the number of information bits divided by the sum of all currently sent modulation symbols:

$$R_{CM}^{ARQ,j} = \frac{L_{info}}{\sum_{i=1}^{j} L_i},\ j \geq 1, \tag{4.2}$$

or expressed with the help of the individual packet coding-/modulation rates

$$R_{CM}^{ARQ,j} = \frac{1}{\sum_{i=1}^{j} \frac{1}{R_{CM}^{FEC,i}}},\ j \geq 1. \tag{4.3}$$

The author likes to emphasis, that the j -th coding-/modulation rate of the ARQ system is a function of the information packet length and the sizes off all packets 1 to j., i.e. it incorporates the already transmitted symbols from previous transmissions. Clearly, the coding-/modulation rates of ARQ systems are monotonically decreasing with the packet number

$$R_{CM}^{ARQ,j+1} < R_{CM}^{ARQ,j},\ j \geq 1,$$

whereas the coding-/modulation rate of the individual packets $R_{CM}^{FEC,j}$ can show any behavior. Specifically, for CE-ARQ systems (constant packet size L)

$$R_{CM}^{FEC,j} = R_{CM}^{FEC} = \frac{L_{info}}{L} \tag{4.4}$$

and the CE-ARQ systems have a hyperbolically decreasing coding-/modulation rate with increasing packet number

$$\begin{aligned} R_{CM}^{ARQ,j} &= \tfrac{1}{j}\tfrac{L_{info}}{L} \\ &= \tfrac{1}{j} \cdot R_{CM}^{FEC},\ j \geq 1. \end{aligned}$$

So far we introduced comparisons with constant conditions on the symbol basis. As mentioned, this approach does not take into account the possibly varying data rate of the systems to be compared. As another approach we could compare two FEC systems on an information rate basis, i.e. in such a way that both systems transmit information at the same rate. To achieve this, despite having a different coding-/ modulation rates, the transmission symbol duration of the systems must be different. A system with a higher coding-/modulation rate can allow for a smaller channel symbol rate in order to match the information rate of a system with a smaller coding-/modulation rate. This different channel symbol rate has two major effects. Firstly, the systems require different bandwidths and secondly, the received signal energy varies (for constant receive signal power). Hence, a system with a higher coding-/modulation rate requires less bandwidth and obtains a higher signal energy than a system with a smaller coding-/modulation rate, if being compared on an information rate basis. In order to describe the changed signal energy , the normalized energy to noise power density $\frac{E_b}{N_0}$ is introduced:

$$\frac{E_b}{N_0} = \frac{1}{R_{CM}} \cdot \frac{E_S}{N_0}. \tag{4.5}$$

For an interpretation of Equation 4.5 consider three FEC system with coding-/modulation rates $\frac{1}{2}$, 1, and 2, respectively. If, for example, the systems are compared at $\frac{E_b}{N_0} = 0\,dB$, the system with coding-/modulation rate 1 operates at an actual channel condition of $\frac{E_S}{N_0} = 0\,dB$. The system coding-/modulation rate $\frac{1}{2}$, however, operates at an channel condition of $\frac{E_S}{N_0} = -3\,dB$ since it has

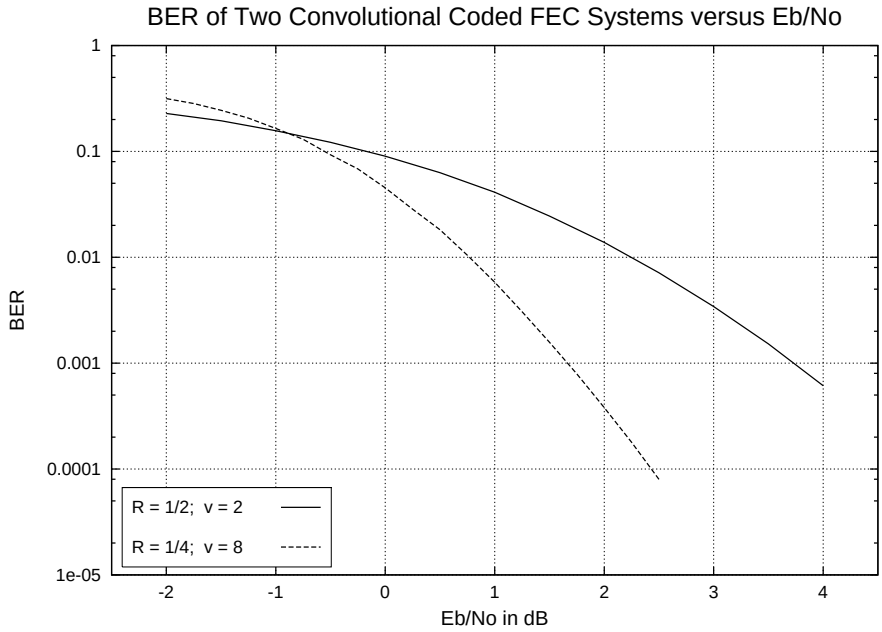

Figure 4.2: BER of Two Convolutional Encoded FEC Systems Using BPSK over AWGN.

only half the channel symbol duration and hence half the channel symbol energy. Similarly, the system with coding-/modulation rate 2 operates at signal-energy-to-noise-power-density ratio of $\frac{E_S}{N_0} = +3\,dB$.

As $\frac{E_S}{N_0}$, $\frac{E_b}{N_0}$ is a dimensionless quantity. In a logarithmic measure the above equation yields

$$\left.\frac{E_b}{N_0}\right|_{dB} = \left.\frac{E_S}{N_0}\right|_{dB} - 10 \cdot log_{10}\left(R_{CM}\right).$$

Hence, performance measures of a system plotted over $\frac{E_b}{N_0}$ instead over $\frac{E_S}{N_0}$ are simply horizontally shifted copies of each other if this measure is *not* a function of the coding-/ modulation rate R_{CM} itself (such as BER and block error rate (BLER) in a FEC system or the number of retransmissions in an ARQ system - the throughput or the data rate of an ARQ system, on the other hand, is itself a function of the coding-/modulation rate and the plot vs. $\frac{E_b}{N_0}$ is not sinply a horizontal shift of the plot vs. $\frac{E_S}{N_0}$).

If we take the two FEC systems with BER vs. $\frac{E_S}{N_0}$ depicted in Figure 4.1, the BER versus $\frac{E_b}{N_0}$ curve of the $R_c = \frac{1}{2}$ system is shifted by 3 dB towards higher signal-to-noise ratios compared to the BER versus $\frac{E_S}{N_0}$ curve. On the other hand, the BER versus $\frac{E_b}{N_0}$ curve of the $R_c = \frac{1}{4}$ system is shifted by 6 dB towards higher signal-to-noise ratios compared to its BER versus $\frac{E_S}{N_0}$ curve. Hence, the two BER versus $\frac{E_b}{N_0}$ curves are 3 dB closer than the corresponding BER versus $\frac{E_S}{N_0}$ curves, putting into perspective the possible coding gain (see Figure 4.2).

> If any performance measures are compared versus $\frac{E_b}{N_0}$, the comparison assumes constant information rate. That means the compared systems have the same information

bit durations and hence possibly different channel symbol rates and accordingly different bandwidth requirements.

Which of the the two approaches is more appropriate depends on the application. In FEC coding theory usually the plots versus $\frac{E_b}{N_0}$ are preferred since it takes the information rate into account and quantifies the coding gain, whereas sometimes in bandwidth limited applications the performance is shown versus $\frac{E_S}{N_0}$. In the next section, we will see that for some performance measures both plots have a distinct meaning and provide different insights into the system. However, in Section 4.4.3 it will be shown that for some ARQ performance measures functional relations with $\frac{E_b}{N_0}$ not necessarily exist.

4.2 Retransmission Request and Rejection Probabilities

4.2.1 Retransmission Request Probability $P(RR_i)$

The successful transmission of a packet over a wireless channel presents a random experiment. As already mentioned, the ARQ principle is to inform the transmitter of the success (via an ACK) or the failure (via a NAK) of the decoding attempt (decision feedback).

In an ARQ system, the event that all decoding events up to the i-th are unsuccessful and hence the i-th retransmission request occurs is denoted as RR_i and the corresponding probability as $P(RR_i)$.

These probabilities are of crucial importance for all performance measures, which will be defined throughout this chapter. For a **memoryless CE-ARQ system**, where the probability that the transmission is rejected is constant, independent of the number of transmission, the retransmission request probability is easy to obtain. If $P(R)$ denotes this constant rejection probability of the individual transmissions, then $P(R)^i$ denotes the probability that the transmissions are rejected i-times. Hence,

$$P(RR_i) = P(R)^i .$$

However, for a **general ARQ system with memory**, the retransmission request probabilities can not be obtained that easy. The main problem is that the probability $P(RR_i)$ not only depends on the current received transmission and hence can not recursively be written as a product. These inherent dependencies severely complicate an analytical evaluation. Also, a numerical simulation is more extensive due to the memory. As an example, if $P(RR_5)$ is to be simulated, only the cases which are rejected at least 4 times contribute to the statistic. Clearly, if retransmission probabilities $P(RR_i)$ with large i are to be simulated, it requires a huge number of simulation runs to gather enough valid experimental runs.

In the following section new probabilities will be defined, which do not suffer from these problems. It is then shown that they can be used to bound the retransmission request probabilities.

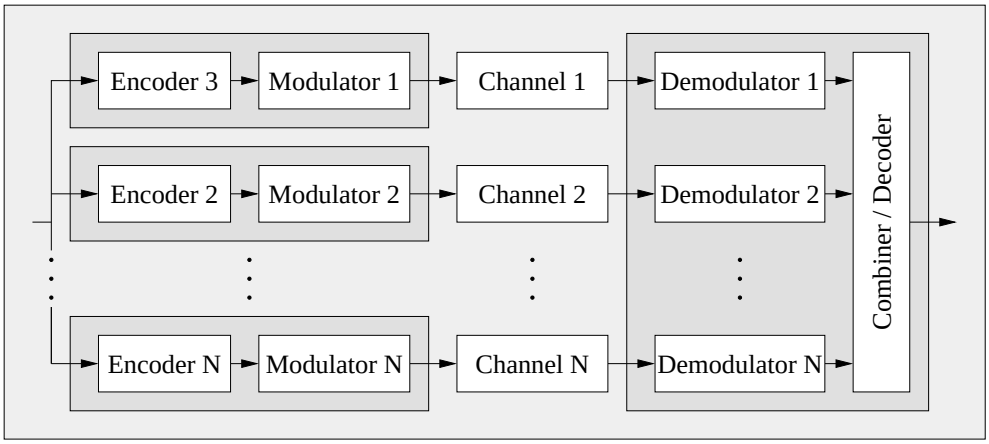

Figure 4.3: Diversity Transmission Scheme.

4.2.2 Rejection Probability $P(R_i)$

In this section related probabilities to the above introduced retransmission request probabilities are defined: the rejection probabilities $P(R_i)$. For their definition, we consider the diversity system depicted in Figure 4.3 consisting of N parallel branches related to the ARQ system of interest. The individual branches identical to the transmission scheme the ARQ system uses at the corresponding transmission, i.e. encoder 1, modulator 1, and demodulator 1 are identical to the ones that the ARQ system uses in its first transmission, encoder 2, modulator 2, and demodulator 2 are identical to the ones that the ARQ system uses in its second transmission, and so on. Additionally, the employed combining method shall be identical in both systems as well as the channel statistics and their possible correlation.

> The event R_N of a rejection of the combined N transmissions of the N diversity scheme of Figure 4.3 is called N rejection and the corresponding probability $P(R_N)$ is the N-th rejection probability.

The rejection probability $P(R_i)$ of the diversity scheme and the retransmission request probability $P(RR_i)$ of the corresponding ARQ system are not identical. To illustrate this consider the combined version of first and second transmission in the ARQ system. A second transmission becomes only necessary if the first one fails. This, however, represents a negative quality predication for the first transmission. For the combined version of transmission one and two in the diversity scheme, on the other hand, no such predication about the transmission no. 1 is made. The same argumentation applies to all rejection probabilities $P(R_i)$ and retransmission request probabilities $P(RR_i)$ with $i > 1$.

In order to reveal the relation of retransmission request events RR_i and the rejection events R_i we compare the $N-1$ and the N diversity schemes and note that they only differ by the N-th branch. Hence, $R_{N-1} \cap R_N$ comprises only tuples of noise sequences, which yield to a failure of the $N-1$ diversity scheme **and** the N diversity scheme (contrary, R_N comprises all ordered

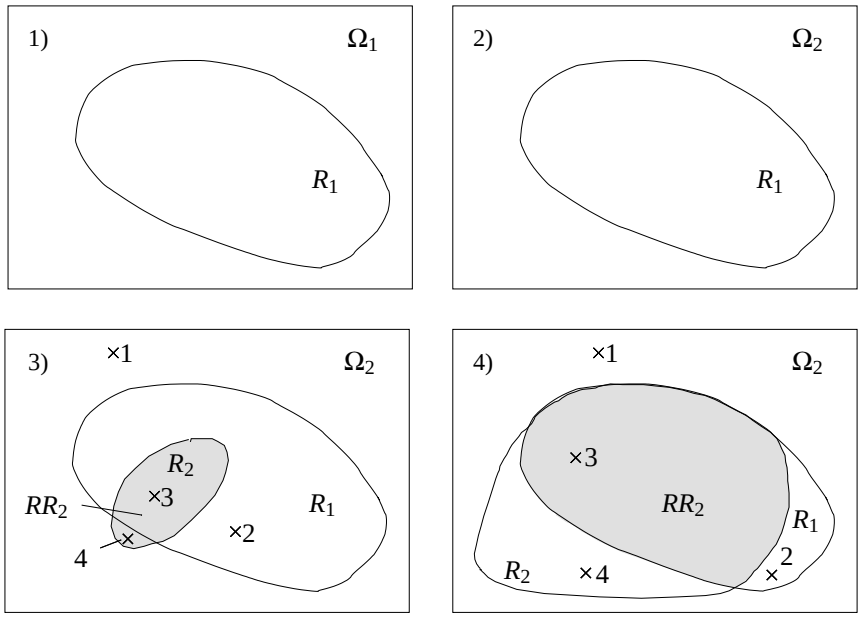

Figure 4.4: Venn Diagram of the Retransmission Request and the Rejection Events.

N-tuples of noise sequences which would fail the decoding process of the N diversity scheme, including those which $(N-1)$-subtuple do not fail the decoding process of the corresponding $(N-1)$-diversity scheme). From the above said follows that event i-th retransmission request RR_i can be expressed with the help of the i rejection events R_j, $j = 1, \ldots, i$ as

$$\begin{aligned} RR_i &= R_1 \cap R_2 \cap \cdots \cap R_i \\ &= \bigcap_{j=1}^{i} R_j \end{aligned}$$

and consequently for the corresponding probabilities

$$P\left(RR_i\right) = P\left(\bigcap_{j=1}^{i} R_j\right).$$

As an illustration for the relation of the retransmission request and the rejection probabilities Figure 4.4 shows several Venn-diagrams, which will be explained in the following.

Figure 4.4 1) shows the space of all possible noise sequences after the first transmission Ω_1 as a rectangular and the subspace R_1 (which is identical to RR_1) of all noise sequences, which yield to a failure of the decoding process[2] .

Figure 4.4 2) is essentially the same as the previous one, but the elements are now the elements

of 1) extended by all possible noise sequences to 2-tuples of noise sequences. Hence Ω_2 represents the space of all possible pairs (ordered 2-tuple) and R_1 comprises all ordered pairs of noise sequences from which the first noise sequence yields a failure of the decoding process in the first branch.

Figure 4.4 3) additionally includes the set of all noise sequence pairs R_2, which yield a failure of the decoding process in the 2-diversity scheme. It also depicts the intersection of R_1 and R_2 as a hatched area. The set of all noise sequence pairs within that area lead to a retransmission request of the corresponding ARQ system.

Figure 4.4 4) depicts also depicts the events of rejection R_1, R_2 and the retransmission request event RR_2 after the second transmission (with two branches) but this time from a different diversity scheme and its corresponding ARQ system.

The first thing we note is that the retransmission request probability, as the probability of the intersection of the individual rejection events

$$\begin{aligned} RR_2 &= R_1 \cap R_2 \\ &\subseteq R_i,\ i = 1, 2 \end{aligned}$$

is always smaller than any of the rejection probabilities:

$$\begin{aligned} P\left(RR_2\right) &= P\left(R_1 \cap R_2\right) \\ &\leq P\left(R_i\right),\ i = 1, 2. \end{aligned}$$

This relation can be generalized to

$$P\left(RR_n\right) \leq P\left(R_i\right),\ i = 1, 2, \ldots, n$$

and hence

$$P\left(RR_n\right) \leq \min_{i=1,\ldots,n} P\left(R_i\right). \tag{4.6}$$

This upper bound on the retransmission request probability was purely based on basic probability calculus. The answer which one of the this probabilities $P\left(R_i\right),\ i = 1, 2, \ldots, n$ is minimal, however, can not be found with mathematics. As it can be seen in Figure 4.4 3) the union $R_1 \cup R_2$ comprises three types of noise sequence tuples.

- Tuples such as no. 2 belong to R_1 but not to R_2, i.e. these noise sequence pairs would result in an erroneous first, but successful second transmission in an ARQ system or a failed transmission of the 1-diversity but a successful transmission with the 2-diversity scheme
- Tuples such as no. 3 belong to both, R_1 and R_2, i.e. the first and second transmission of the ARQ system fail as well as the transmissions of the 1 and 2-diversity schemes.

[2] A linear code is assumed, where the event wrong decoding is purely dependent on the noise sequence.

If only noise tuples of these types would exist it is obvious that

$$P(R_2) \leq P(R_1).$$

However, there are also noise sequence pairs in the union of the following type

- Tuples such as no. 4 which do not belong to R_1 but to R_2. For these samples, the behavior of the ARQ system and the diversity scheme differ. After the first transmission, the ARQ system would successfully decode the transmission and the transmission process is successfully finished. The same applies to the 1-diversity scheme, but the 2-diversity scheme fails.

This represents the cases where the combined version of a qualitatively good first transmission (good enough to be successfully decoded on its own) is corrupted by a bad second transmission. Whether $P(R_1)$ or $P(R_2)$ is smaller depends on the likelihood of this performance degradation , i. e. $P(R_2 \setminus R_1)$, in comparison to the performance improvement by the combining, i.e. $P(R_1 \setminus R_2)$. Figure 4.4 3) graphically illustrates a case where the performance improvement outbalances the performance degradation, whereas Figure 4.4 4) illustrates the opposite case. However, for any reasonable combing system the performance should improve on average (this does not exclude noise realizations, where the combining worsens the result - it simply means that these cases are rare in comparison to the cases, leading to an improvements), i.e.

$$P(R_1) \geq P(R_2) \geq \cdots \geq P(R_n) \geq \cdots \tag{4.7}$$

and the tightest upper bound in Inequation 4.6 is the rejection probability with the highest index. Hence,

$$P(RR_n) \leq P(R_n)$$

will be used in the remainder of this work as upper bound.

In order to obtain a lower bound, we make use of Bayes Rule

$$P(A|B) \cdot P(B) = P(A \cap B)$$

and write the retransmission request probability as

$$\begin{aligned} P(RR_2) &= P(R_1 \cap R_2) \\ &= P(R_2|R_1) \cdot P(R_1). \end{aligned}$$

The conditioned probability $P(R_2|R_1)$ is the probability of a failure of the 2-diversity scheme but only over the noise sequence tuples whose first noise sequence leads to a failure in the 1-diversity scheme. Independent of the quality of the combining, this probability is larger than it would be if calculated over all possible noise sequence tuples, i.e. Ω_2. Hence,

$$P(R_2|R_1) \geq P(R_2)$$

and

$$P\left(RR_2\right) \geq P\left(R_2\right) \cdot P\left(R_1\right).$$

For a generalization to the n-th retransmission request probability $P\left(RR_n\right)$ we recursively use Bayes Rule:

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} R_i\right) &= P\left(R_n \left| \bigcap_{i=1}^{n-1} R_i \right.\right) \cdot P\left(\bigcap_{i=1}^{n-1} R_i\right) \\ &= P\left(R_n \left| \bigcap_{i=1}^{n-1} R_i \right.\right) \cdot P\left(R_{n-1} \left| \bigcap_{i=1}^{n-2} R_i \right.\right) \cdot P\left(\bigcap_{i=1}^{n-2} R_i\right) \\ &\vdots \\ &= P\left(R_n \left| \bigcap_{i=1}^{n-1} R_i \right.\right) \cdot P\left(R_{n-1} \left| \bigcap_{i=1}^{n-2} R_i \right.\right) \cdots P\left(R_2 \left| R_1 \right.\right) \cdot P\left(R_1\right) \\ &= P\left(R_1\right) \cdot \prod_{i=0}^{n-2} P\left(R_{n-i} \left| \bigcap_{m=1}^{n-i-1} R_m \right.\right). \end{aligned} \tag{4.8}$$

Just as in the argumentation of $P\left(R_2 \left| R_1 \right.\right) \geq P\left(R_2\right)$, we can use the same argumentation to obtain a general lower bound for all conditioned probabilities in Equation 4.8:

$$P\left(R_{n-i} \left| \bigcap_{m=1}^{n-i-1} R_m \right.\right) \geq P\left(R_{n-i}\right).$$

Replacing the conditional probabilities in Equation 4.8 with their corresponding lower bounds, reordering, and applying an index transformation yields

$$\begin{aligned} P\left(\bigcap_{i=1}^{n} R_i\right) &\geq P\left(R_1\right) \cdot \prod_{i=0}^{n-2} P\left(R_{n-i}\right) \\ &= P\left(R_1\right) \cdot \prod_{i=0}^{n-2} P\left(R_{2+i}\right) \\ &\vdots \\ &= P\left(R_1\right) \cdot \prod_{i=2}^{n} P\left(R_i\right) \\ &= \prod_{i=1}^{n} P\left(R_i\right). \end{aligned}$$

Hence,

$$\prod_{j=1}^{n} P\left(R_j\right) \leq P\left(\bigcap_{j=1}^{n} R_j\right) = P\left(RR_n\right) \leq P\left(R_n\right). \tag{4.9}$$

According to Inequation 4.9, the rejection probabilities of the diversity scheme of Figure 4.3 can be used to bound the more difficult to obtain retransmission request probabilities of the corresponding ARQ system. The author likes to emphasize, that for the derivation of the bound, no

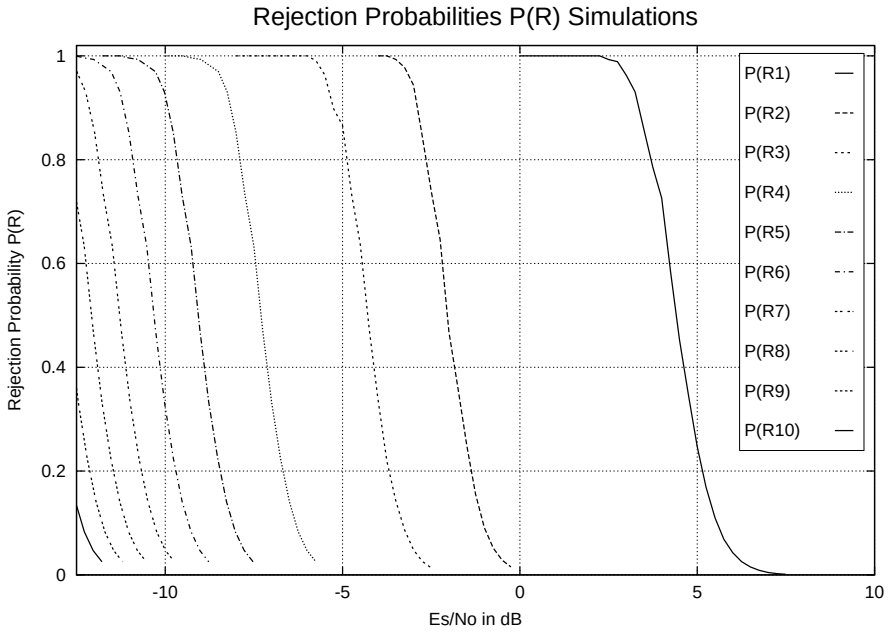

Figure 4.5: Rejection Probabilities $P(R_j)$ of a Punctured, Rate $R_C = \frac{1}{3}$ Convolutional Encoded ($v = 8$) VE-MARQ System using BPSK over AWGN versus $\frac{E_S}{N_0}$.

assumptions about the channel and especially about the correlation of the noise sequences of the different transmission were made. As long as the diversity scheme and the corresponding ARQ scheme have the same individual channel statistics and the same correlation, Inequation 4.9 holds.

4.2.3 Tightness of the Retransmission Request Bound

The tightness of the bound 4.9 heavily determines the tightness of all other bounds, which will be derived in the course of this chapter. We therefore pay some attention to this issue.

The general behavior of the rejection probability is that all $P(R_j) = 1$ for $\frac{E_S}{N_0}$ values below certain values and $P(R_j) = 0$ for $\frac{E_S}{N_0}$ values above certain values[3]. Also, as already mentioned the higher the index j, the smaller are the corresponding rejection probabilities for the same channel condition. Figure 4.5 depicts simulations of the rejection probabilities $P(R_j)$ of j-diversity schemes, $j = 1, \ldots, 10$ constructed after a VE-MARQ system[4] using a punctured convolutional code with rate $R_C = \frac{1}{3}$ and constraint length $v = 8$ and BPSK over a AWGN channel. The mentioned behavior of the rejection probabilities is clearly visible. For each $\frac{E_S}{N_0}$ value, all inequations in Inequation 4.7 are satisfied and also the region were the rejection probabilities effectively vary between 1 and 0 is limited to a 4 dB region.

[3] The rejection probabilities are never absolutely equal to zero or one. Within that section, equal to zero or one refers to a proability which is almost equal to zero or one, i.e. that it can not be distinguished from zeo or one in a linear plot.

[4] In order to avoid the rather cumbersome description "the rejection probabilities of the N-diversity scheme constructed after a certain ARQ scheme" we will use the terminology "rejection probabilities of an ARQ system", keeping in mind that we mean the rejection probabilities of the corresponding diversity schemes.

For the first retransmission request probability $P(RR_1)$, Bound 4.9 yields

$$P(R_1) \leq P(RR_1) \leq P(R_1)$$

and it follows that the first rejection and the first retransmission request probability are always identical

$$P(R_1) = P(RR_1).$$

For the second retransmission request probability follows

$$P(R_1) \cdot P(R_2) \leq P(RR_2) \leq P(R_2)$$

and the tightness of the upper and lower bound depends on the actual shape of the rejection probabilities $P(R_1)$ and $P(R_2)$. For our example, we can obtain from Figure 4.5 that $P(R_1) = 1$ for all $\frac{E_S}{N_0}$ values where $P(R_2) \neq 0$ and hence, $P(R_2) = P(RR_2)$ over the complete $\frac{E_S}{N_0}$ range. More general, if we rewrite Inequation 4.9 as

$$\left[\prod_{j=1}^{n-1} P(R_j)\right] \cdot P(R_n) \leq P(RR_n) \leq P(R_n)$$

we can state that following rule:

> The n-th retransmission request probability $P(RR_n)$ is identical to the n-th rejection probability $P(R_n)$ for the $\frac{E_S}{N_0}$ region where the rejection probability $P(RR_{n-1})$ with the next lower index $n-1$ is one (due to Inequation 4.7 all $P(R_j)$ with $j < n-1$ are also be 1).

Hence, by looking at Figure 4.5 it becomes obvious that $P(RR_3) = P(R_3)$ for $\frac{E_S}{N_0} < -4\,dB$. With an increasing index, the rejection probabilities start noticeably to overlap and we expect the bound to diverge.

Figure 4.6 depicts the resulting bounds on the retransmission probabilities $P(RR_j)$ derived from the rejection probabilities shown in Figure 4.5. As argued, for $P(RR_1)$ and $P(RR_2)$ the upper and lower bounds are identical, whereas the bounds of $P(RR_3)$ differ (although hardly noticeable) for $\frac{E_S}{N_0} > -4\,dB$. For higher indices, the bounds start to deviate, although the relative difference is quite small.

The good quality of the bounds for the presented example is a result of the sharp decrease of the $P(R_j)$ and the relatively large separation between the individual rejection probabilities.

For the small extension of the region where the rejection probabilities vary between 0 and 1 the following things are accountable

- The quality of the code. With an improved code not only the $P(R_j)$ are shifted towards lower $\frac{E_S}{N_0}$ values, but also the gradient of the curves increase.
- The information block length. With an increased block length, the block error rate increases

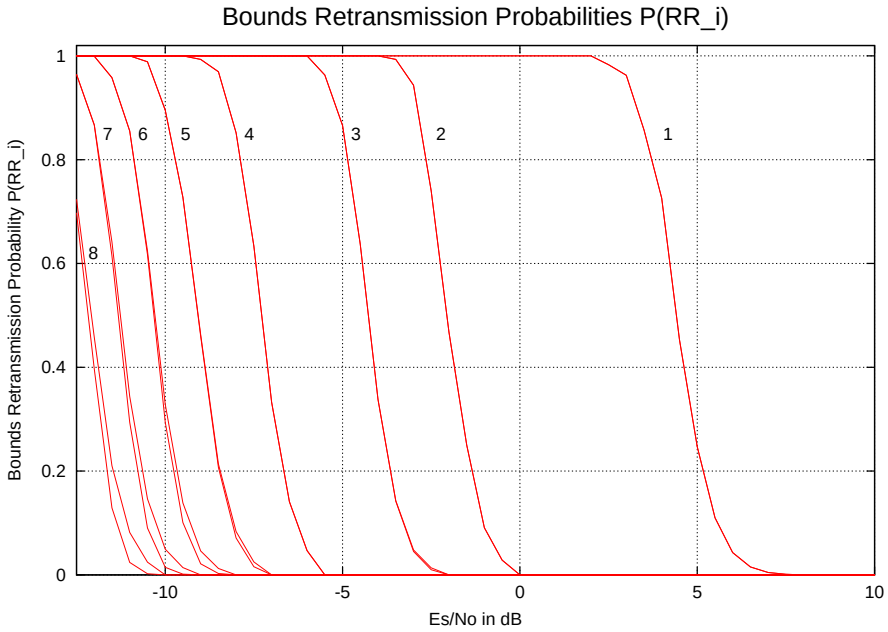

Figure 4.6: Bounds of the Retransmission Request Probabilities $P\left(RR_j\right)$ for the VE-MARQ System with Rejection Probabilities Depicted in Figure 4.5.

even if the bit error rate remains constant. As a simple illustration regard the bit errors independently distributed thoughout the information word. Then, the block error probability $BLER_N$ of a block with N information bits is $BLER_N = 1 - (1 - BER)^N$ if the bit error probability is BER. Although for a practical code this independent distribution is not given, still a similar exponential behavior can be noted.

- The channel. The less the channel varies its statistics, the smaller the region of interest becomes. Of all channel models presented in Chapter 3, the AWGN channel shows the least, whereas the constant Rayleigh channel has the largest statistical variations.

For the separation between the individual rejection probabilities, again, mainly the following things are responsible

- The combining. The better the code, which is constructed with the retransmissions, the larger the separation is.

- The channel. The smaller the variations of the channel, the larger the separations are. Hence, the AWGN has the largest separations, whereas the constant Rayleigh has the smallest.

From the above stated, it becomes clear that the presented example in Figures 4.5 and 4.6 satisfies most criteria for tight bounds. The AWGN channel, the used block length of 280 information bits, the high constraint length convolutional code, as well as the punctured combining scheme contribute to tight bounds even for high retransmission request probability numbers.

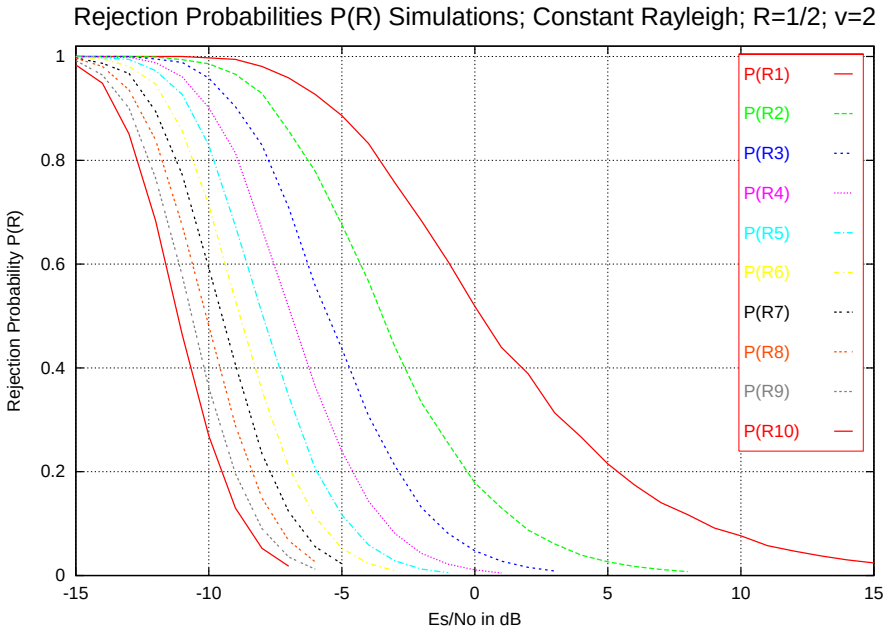

Figure 4.7: Rejection Probabilities of CE-ARQ using a Rate $R_C = \frac{1}{2}$ Convolutional Code ($v = 2$) over a Constant Rayleigh Channel versus the Average $\frac{E_S}{N_0}$.

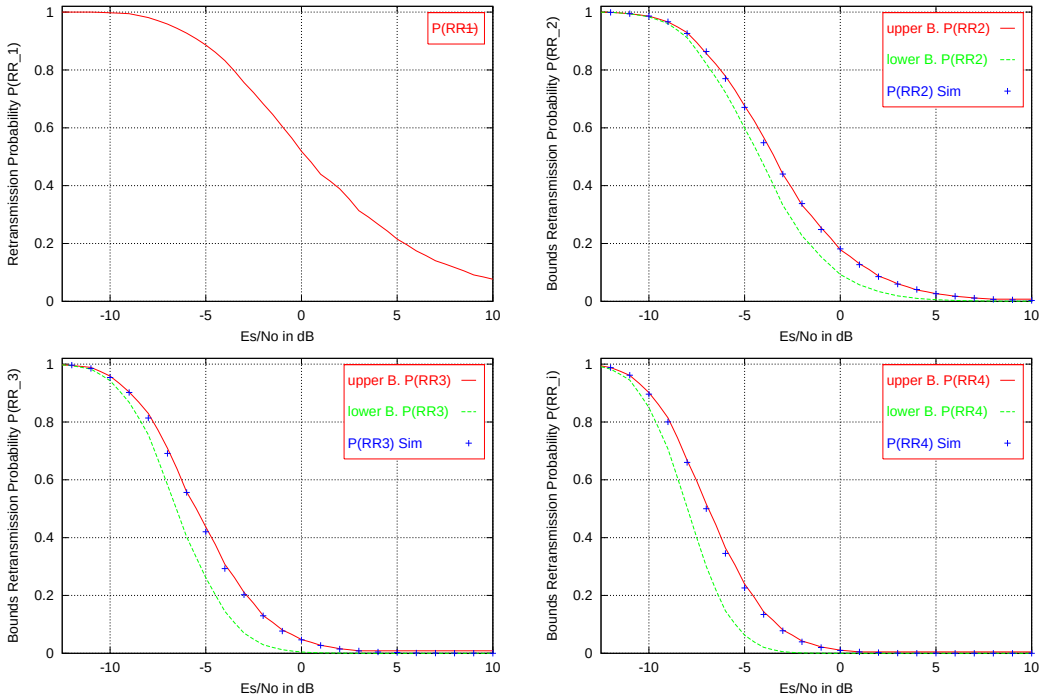

Figure 4.8: Bounds of the Retransmission Request Probabilities $P\left(RR_j\right)$ for the VE-MARQ System with Rejection Probabilities Depicted in Figure 4.7.

As a counter example, we use the severe constant Rayleigh channel, a bad code (rate $\frac{1}{2}$, constraint length 2 convolutional code), a short block length of 100 information bits, and as retransmissions repetitions, which will be combined via maximum ratio combining (MRC; see Section 5.1.4). Figure 4.7 shows the resulting first 10 rejection probabilities. Due to the channel, the code, and the small block length, the rejection probabilities decrease much slower. Again due to the channel, they are also not simple horizontal shifts of each other. As a result, there are no $\frac{E_S}{N_0}$ regions where consecutive rejection probabilities do not overlap. Consequently, we expect the corresponding bound for the retransmission request probability to be much looser than in the AWGN channel. Figure 4.8 depict the resulting bounds on $P(RR_j)$ in conjunction with simulations of these probabilities. At first it becomes apparent, that, as expected, the bounds are much looser. For all depicted $P(RR_j)$ they differ up to 2 dB. A more interesting fact is revealed by the comparison of the simulation result with the bounds, namely that the retransmission request probabilities $P(RR_j)$ are approximately equal to their upper bounds (the rejection probabilities $P(R_j)$). This is a result of the optimum combining method, which yields to a almost zero probability that an excellent first transmission (at least good enough to be successfully decoded by the ARQ system) is corrupted by a bad second transmission in such a way that the combined version fails the coding attempt. That is, the probability of instances such as point 4 in Figure 4.4 3) on Page 54 is vanishing small. In general, the optimum combining leads to a vanishing probability that an additional transmission leads to an unsuccessfull decoding if the previous combined transmissions are decoded successfully.

4.2.4 Interpretation of $P(R_j)$ vs. $\frac{E_S}{N_0}$ Plot

The shape of the rejection probabilities and their mutual relation depends on the ARQ type (CE or VE), the decoder strategy (with or without memory), and on the channel model. As an example, Figure 4.5 shows the individual rejection probabilities $P(R_j), j = 1, \ldots, 10$ of a CE-MARQ system versus $\frac{E_S}{N_0}$.

A $P(R_j)$ vs. $\frac{E_S}{N_0}$ plot reveals if the granularity of the redundancy delivered by the ARQ system is fine enough, so that the system is not overdesigned at some $\frac{E_S}{N_0}$ values. In order to satisfy this criteria, the individual rejection probabilities should overlap for the most part. In Figure 4.5 it is apparent that for that specific system this is only true for $\frac{E_S}{N_0} < -6\,dB$. As already mentioned, $P(R_1)$ and $P(R_2)$ do not overlap at all between 0 and 2.5 dB. This means that the redundancy provided by the second transmission is too much for that $\frac{E_S}{N_0}$ region, resulting in an overdesigned system with unnecessary low information rate. Hence, we can state the following rule:

> Abscissa values in a $P(R_j)$ vs. $\frac{E_S}{N_0}$ plot, where all rejection probabilities are either zero or one reveal signal-energy-to-noise-power-density ratios in which the ARQ system is overdesigned.

In the following we will discover other behaviors of performance measures, which are symptoms for a bad system design. Also, consequences of this overdesign will be discussed throughout this chapter.

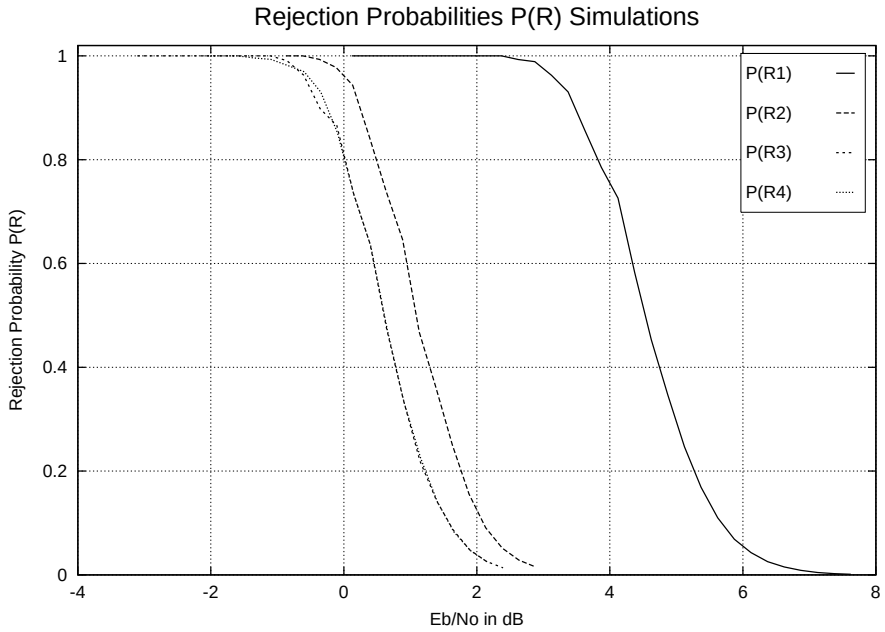

Figure 4.9: Rejection Probabilities of a Punctured Rate $R_C = \frac{1}{3}$ Convolutional Encoded ($v = 8$) VE-MARQ System using BPSK over AWGN versus $\frac{E_b}{N_0}$.

4.2.5 Interpretation of $P(R_j)$ vs. $\frac{E_b}{N_0}$ Plot

As discussed in the previous sections of this chapter, the rejection probabilities $P(R_j)$ could also be plotted against the normalized energy to noise power density ratio $\frac{E_b}{N_0}$. By doing so, the loss in information rate with the further transmissions is taken into account. As argued in Section 4.1, when plotting the rejection probabilities of an ARQ system versus $\frac{E_b}{N_0}$ the the ARQ coding-/modulation rate $R_{CM}^{ARQ,j}$ (see Equation 4.3) should be used to convert from $\frac{E_S}{N_0}$ to $\frac{E_b}{N_0}$.

Figure 4.9 depicts the same rejection probabilities $P(R_j)$ as Figure 4.5, but this time versus the corresponding $\frac{E_b}{N_0}$. It can be seen how the gaps between $P(R_1)$, $P(R_2)$, and $P(R_3)$ - as expected - are downsized. The more dramtic revealment of Figure 4.9 is that $P(R_3)$, $P(R_4)$, and all further probabilities $P(R_j)$, $j \geq 5$ (not shown in Figure 4.9) are identical. This immediately implies that the $4th$ and all succeeding transmissions are simply repetitions of the code constructed with the first three transmissions. In coding theoretical terms, the system does not improve with the 4-th and all following transmissions. Hence:

> The $P(R_j)$ vs. $\frac{E_b}{N_0}$ plot provides a coding theoretical insight into an ARQ system. It indicates the quality of the code constructions with progressional transmissions.

As another example lets have a look at the rejection probabilities $P(R_j)$ vs. $\frac{E_b}{N_0}$ of a memoryless CE-ARQ system depicted in Figure 4.10. The steadily decreasing coding-/modulation rate with the increased number of transmissions and the constant retransmission request probability results in increasing shifts of the $P(R_j)$ toward higher $\frac{E_b}{N_0}$ values. This immediately demonstrates the

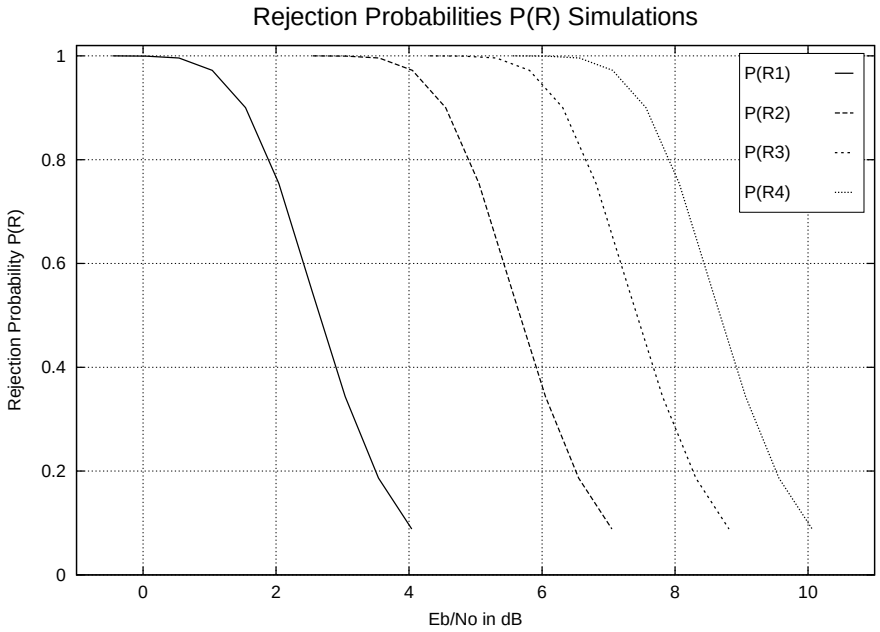

Figure 4.10: Rejection Probabilities of a memoryless CE-ARQ System versus $\frac{E_b}{N_0}$.

bad performance of this ARQ class in coding theoretical terms.

4.3 Number of Transmissions $\mathbf{n}_{trans}$

The principle of ARQ systems is to request additional information if the decoding of a packet was not successful. Depending on the channel and the employed ARQ system, a varying amount of transmissions are necessary until the information is delivered to the sink without errors[5]. The channel condition is usually described by a statistical property, such as the average receive signal-energy-to-noise-power-density ratio $\overline{\frac{E_S}{N_0}}$, and the transmission over a channel represents a statistical experiment. As a consequence, the number of required transmissions $\mathbf{n}_{trans}$ is a random variable. In Section 4.3.1 its probability distribution $P\left(\mathbf{n}_{trans} = x\right)$ and cumulative distribution functions $P\left(\mathbf{n}_{trans} < x\right)$ are derived and bounded with the previously introduced rejection probabilities. They can be used to determine the maximum number of required transmissions with a given probability. Section 4.3.2 investigates its expected value $\overline{n_{trans}} = E\left\{\mathbf{n}_{trans}\right\}$ and also uses the rejection probabilities to derive upper and lower bounds.

4.3.1 Probability Distribution and Cumulative Distribution Functions

The maximum number of transmissions, which occurs during the transmission process of an information packet in an ARQ system, is of crucial importance. It is closely related to different maximum delays and memory requirements, which will be treated later in this chapter. However,

[5] Actually without detected errors.

due to the statistical nature of $\mathbf{n}_{trans}$, there is no absolute maximum number of transmissions. Instead, with each number of transmission n, a certain probability $P(\mathbf{n}_{trans} = n)$ is associated and the probability that we have less than n transmissions is given by the cumulative distribution function $P(\mathbf{n}_{trans} < n)$.

Memoryless CE-ARQ Systems

The feature of memoryless CE-ARQ systems is their *constant* retransmission request probability $P(RR_j) = const$ (see Section 2.1). Since in these systems only the last received packet is used for decoding and all retransmissions are repetitions, the retransmission request probability is constant and equal to the also constant rejection probability $P(R)$.

The probability that we have exactly one transmission is then $1-P(R)$, and for the probability of two transmissions the packet needs to be rejected only once, i.e. $P(R) \cdot (1 - P(R))$, and so on. Consequently, the random variable number of transmissions $\mathbf{n}_{trans}$ has the following probability distribution:

$$\begin{aligned} P(\mathbf{n}_{trans} = 1) &= 1 - P(R) \\ P(\mathbf{n}_{trans} = 2) &= P(R) \cdot (1 - P(R)) \\ P(\mathbf{n}_{trans} = 3) &= P(R)^2 \cdot (1 - P(R)) \\ &\vdots \\ P(\mathbf{n}_{trans} = n) &= P(R)^{n-1} \cdot (1 - P(R)) \\ &\vdots \end{aligned} \tag{4.10}$$

Hence, the cumulative distribution function is

$$P(\mathbf{n}_{trans} < n) = \begin{cases} 0 & , n = 1 \\ \sum_{k=1}^{n-1} P(R)^{k-1} \cdot (1 - P(R)) & , n > 1 \end{cases}$$

and in Appendix C.3.1 it is shown that this is equal to

$$P(\mathbf{n}_{trans} < n) = \begin{cases} 0 & , n = 1 \\ 1 - P(R)^{n-1} & , n > 1. \end{cases} \tag{4.11}$$

Due to the - as a general rule - not analytically available rejection probability $P(R)$, the inversion of Equation 4.11 is also not analytically possible. Nevertheless, the integer value n_{max}, so that the number of required number of transmissions $\mathbf{n}_{trans}$, is smaller or equal to n with a probability P (or an error probability $P = 1 - P_E$) can be obtained numerically for a given rejection probability

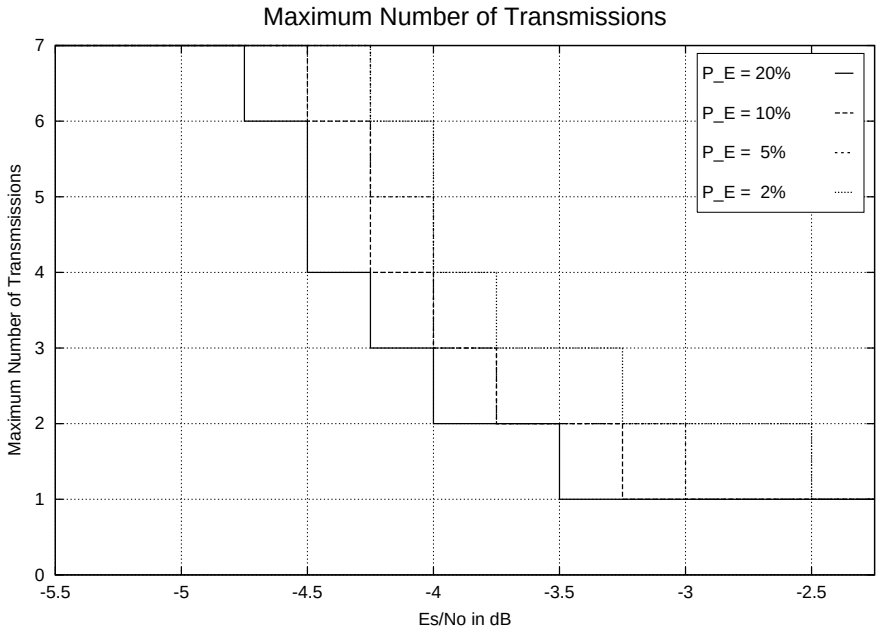

Figure 4.11: Number of Maximal Required Transmissions with Several Reliability Probabilities P for the memoryless CE-ARQ System with the Rejection Probability $P(R) = P(R_3)$ of Figure 4.5.

$P(R)$: From Equation 4.11 it follows

$$\begin{aligned} P(\mathbf{n}_{trans} \leq n) &= P(\mathbf{n}_{trans} < n+1) \\ &= 1 - P(R)^n . \end{aligned} \tag{4.12}$$

Hence, n_{max} is the smallest possible integer satisfying the following inequations:

$$1 - P(R)^{n_{max}} \geq P$$

or

$$P(R)^{n_{max}} \leq P_E.$$

As an example, we consider the memoryless CE-ARQ system with the constant rejection probability $P(R) = P(R_3)$ of Figure 4.5. Figure 4.11 shows the results for several reliabilities. As it can be seen, the tougher the requisitions on the reliability P are, the larger is the maximum number of transmission to expect. At $-4.2\,dB$, for example, we have with a probability of 80% less than 4 transmissions, with a probability of 90% less than 5 transmissions, with a probability of 95% less than 6 transmissions, and with a probability of 98% less than 7 transmissions.

VE-ARQ and MARQ Systems

Contrary to memoryless CE-ARQ systems, VE-ARQ systems have a varying error correction capability and there is no unique rejection probability for the first transmission and all retransmissions. Also, if the schemes incorporate memory, the retransmission probabilities are not necessarily independent. Hence, for these systems the above derivation is not valid.

Let R_1 denote the event rejection of the first packet, R_2 denotes the event rejection of the second packet, and so on. Then, $P\left(\bigcap_{k=1}^{n} R_k\right)$ denotes the probability that all packets up to the n-th transmission are rejected. The random variable $\mathbf{n}_{trans}$ takes the value 1 when the packet is not rejected at its first decoding attempt. The probability for that event is $1 - P(R_1)$. In general, $\mathbf{n}_{trans}$ takes the value n when the packet is rejected $n-1$ times, but not rejected n times. The corresponding probability is

$$P\left(\bigcap_{k=1}^{n-1} R_k\right) \cdot \left(1 - P\left(\bigcap_{k=1}^{n} R_k \middle| \bigcap_{i=k}^{n-1} R_k\right)\right)$$

and as a consequence $\mathbf{n}_{trans}$ has the following probability distribution

$$\begin{aligned}
P(\mathbf{n}_{trans}=1) &= 1 - P(R_1) \\
P(\mathbf{n}_{trans}=2) &= P(R_1)\cdot(1 - P(R_1 \cap R_2 | R_1)) \\
P(\mathbf{n}_{trans}=3) &= P(R_1 \cap R_2)\cdot(1 - P(R_1 \cap R_2 \cap R_3 | R_1 \cap R_2)) \\
&\vdots \\
P(\mathbf{n}_{trans}=n) &= P\left(\textstyle\bigcap_{k=1}^{n-1} R_k\right)\cdot\left(1 - P\left(\textstyle\bigcap_{k=1}^{n} R_k | \bigcap_{k=1}^{n-1} R_k\right)\right) \\
&\vdots
\end{aligned} \tag{4.13}$$

and cumulative distribution for $n > 1$ ($P(\mathbf{n}_{trans} < 1) = 0$)

$$\begin{aligned}
P(\mathbf{n}_{trans} < n) &= \textstyle\sum_{k=1}^{n-1}\left[P\left(\bigcap_{i=1}^{k-1} R_i\right)\cdot\left(1 - P\left(\bigcap_{i=1}^{k} R_i \middle| \bigcap_{i=1}^{k-1} R_i\right)\right)\right] \\
&= \textstyle\sum_{k=1}^{n-1}\left[P\left(\bigcap_{i=1}^{k-1} R_i\right) - P\left(\bigcap_{i=1}^{k} R_i\right)\right].
\end{aligned}$$

In Appendix C.3.2 is is shown that this is equal to

$$P(\mathbf{n}_{trans} < n) = 1 - P\left(\bigcap_{k=1}^{n-1} R_k\right),\ n > 1 \tag{4.14}$$

and using the bounds of Inequation 4.9, the cumulative distribution of the number of transmis-

sions of any MARQ system can be bounded for $n > 1$ as

$$1 - P\left(R_{n-1}\right) \leq P\left(\mathbf{n}_{trans} < n\right) \leq 1 - \prod_{k=1}^{n-1} P\left(R_k\right).$$

Again, the integer value n_{max}, so that the number of required number of transmissions $\mathbf{n}_{trans}$, is smaller or equal to n with a probability P (or an error probability $P = 1 - P_E$) can be obtained numerically for a given rejection probability $P\left(R\right)$: From Equation 4.14 follows

$$\begin{aligned} P\left(\mathbf{n}_{trans} \leq n\right) &= P\left(\mathbf{n}_{trans} < n+1\right) \\ &= 1 - P\left(\textstyle\bigcap_{k=1}^{n} R_k\right). \end{aligned}$$

Hence, n_{max} is the smallest possible integer satisfying the following inequations:

$$1 - P\left(\bigcap_{k=1}^{n_{max}} R_k\right) \geq P \tag{4.15}$$

or

$$P\left(\bigcap_{k=1}^{n_{max}} R_k\right) \leq P_E. \tag{4.16}$$

In order to derive an upper bound $n_{max}^{(upper)}$ (lower bound $n_{max}^{(lower)}$) for n_{max}, the bounds of Inequation 4.9 are used in conjunction with the two preceding inequations. Therefore, the intersection probability in Inequations 4.15 and 4.16 must be replaced with its lower bound (upper bound). Hence, the smallest possible integer $n_{max}^{(upper)}$ satisfying the inequations

$$1 - P\left(\prod_{i=1}^{n_{max}^{(upper)}} R\right) \geq P$$

or

$$P\left(\prod_{i=1}^{n_{max}^{(upper)}} R_i\right) \leq P_E$$

is an upper bound to the maximum number of required transmission with a minimum probability P and a maximum error probability P_E. Analoge, the smallest possible integer $n_{max}^{(lower)}$ satisfying the inequations

$$1 - P\left(R_{n_{max}^{(lower)}}\right) \geq P$$

or

$$P\left(R_{n_{max}^{(lower)}}\right) \leq P_E$$

is an lower bound to the maximum number of required transmission with a minimum probability

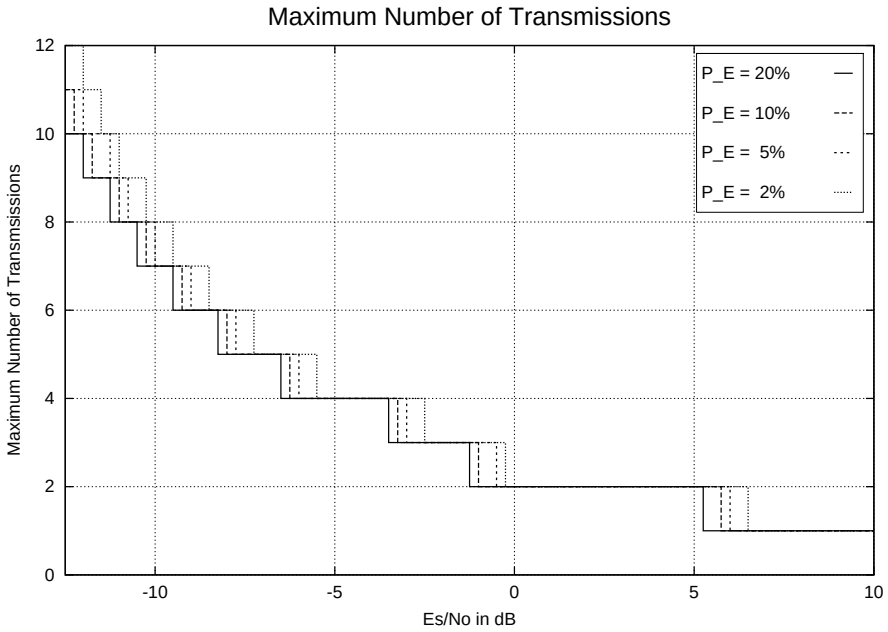

Figure 4.12: Bounds to the Number of Maximal Required Transmissions with Several Reliability Probabilities P for the VE-MARQ System with the Rejection Probabilities $P(R_j)$ of Figure 4.5.

P and a maximum error probability P_E. Figure 4.12 shows the upper bounds to the maximal required transmissions $n_{max}^{(upper)}$ for the VE-MARQ system with the rejection probabilities depicted in Figure 4.5 with several degrees of reliability. The lower bounds $n_{max}^{(lower)}$ for this system are indistinguishable from its corresponding upper bounds and hence not shown in that figure.

4.3.2 Average Number of Transmissions $\overline{n_{trans}}$

In previous subsection, the probability distribution and the cumulative distribution of the number of transmissions $\mathbf{n}_{trans}$ were investigated. Its expected value $\overline{n_{trans}} = E\{\mathbf{n}_{trans}\}$ is also an important performance measure since it is closely related to the throughput and the delays of an ARQ system. In the following the average number of transmissions in dependency of the different rejection probabilities $P(R_j)$ for the various ARQ types are derived.

Memoryless CE-ARQ Systems

Equation 4.10 states the probability distribution for this ARQ class. In Appendix C.3.3 on Page 188 it is shown that the expected value of this distribution is

$$\overline{n_{trans}} = E\{\mathbf{n}_{trans}\} = \frac{1}{1 - P(R)}. \tag{4.17}$$

Equation 4.17 reveals a simple relation between the rejection probability and the on average required number of transmissions. As an example, a 50 % chance of failure results in an average of

2 transmissions and a 90 % failure results on average in 10 transmissions.

In Section 4.2.3 is was mentioned that the $\frac{E_S}{N_0}$-region in which the rejection probability $P(R)$ increases from 0 to 1 as one moves towards worse channel conditions is only a few dB wide. As a consequence, the average number of transmission increases within this small region from 1 to infinity. Equation 4.17 reflects the bad coding theoretical performance of memoryless CE-ARQ systems in terms of the required transmissions.

VE-ARQ and MARQ Systems

The probability distribution for this ARQ types were derived preceeding section (Equation 4.13). In Appendix C.3.4 it is shown that the average number of transmissions for a packet to be accepted by the receiver in dependency of the individual retransmission request probabilities is

$$\begin{aligned} \overline{n_{trans}} &= E\{\mathbf{n}_{trans}\} \\ &= 1 + P(R_1) + P(R_1 \cap R_2) + \cdots + P(R_1 \cap R_2 \cap \cdots \cap R_k) + \cdots \\ &= 1 + \textstyle\sum_{k=1}^{\infty} P\left(\bigcap_{i=1}^{k} R_i\right) \end{aligned} \tag{4.18}$$

However, the probabilities of the joint events $\bigcap_{i=1}^{k} R_i$ are for MARQ systems, due to their statistical dependency, difficult to obtain.

On the other hand, if no memory is used, the events R_i are independent, the probability of the joint event is equal to the product of the individual event probabilities and Equation 4.18 reduces to

$$\overline{n_{trans}} = 1 + \sum_{k=1}^{\infty} \left[\prod_{i=1}^{k} P(R_i) \right],$$

which is a generalized version of Equation 4.17, also applicable to memoryless VE-ARQ systems.

If, however, the events R_i are not independent, Equation 4.18 can again be bounded with the help of rejection probabilities $P(R_i)$. Therefore, Inequation 4.9 is used to bound the intersection probabilities, yielding

$$1 + \sum_{k=1}^{\infty} \left[\prod_{i=1}^{k} P(R_i) \right] \leq \overline{n_{trans}} \leq 1 + \sum_{k=1}^{\infty} P(R_k) \ . \tag{4.19}$$

The tightness of these bounds depends on the the individual retransmission probabilities $P(R_i)$. If the $\frac{E_S}{N_0}$- regions, where $P(R_{i+1})$ is effectively grater than zero and $P(R_i)$ is effectively smaller than 1, do not overlap, the bounds will be very tight.

Figure 4.13 depicts the simulated average number of transmissions $\overline{n_{trans}}$ of the system with the rejection probabilities $P(R_j)$ shown in Figure 4.5 on Page 58 together with the upper and lower bound derived according to Equation 4.19 from the simulated rejection probabilities.

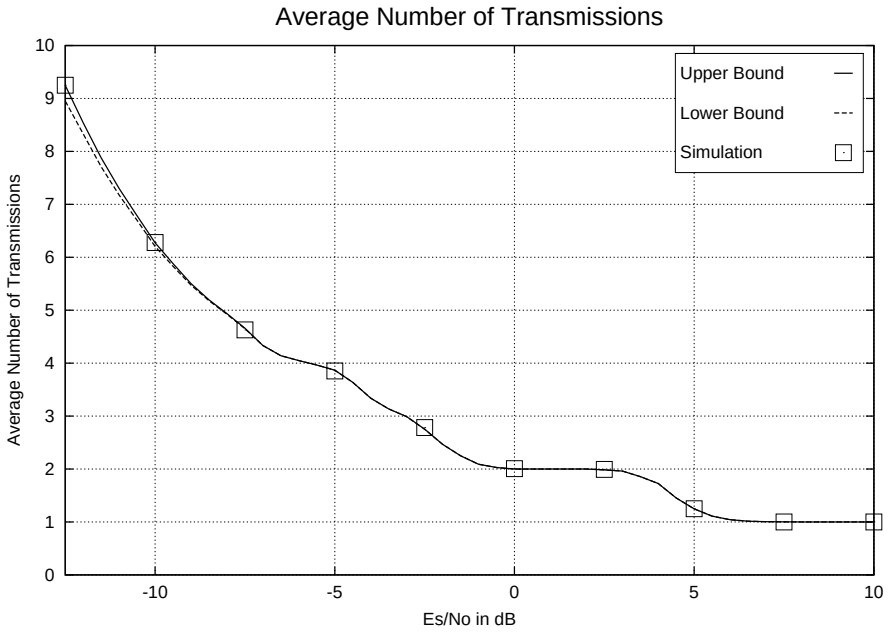

Figure 4.13: Bounds on the Average Number of Transmissions of the ARQ System with the Retransmission Probabilities Depicted in Figure 4.5.

The first thing which becomes apparent is the tightness of the bounds. The bound start slowly deviating if the $\frac{E_S}{N_0}$ falls below $-10\ dB$. As mentioned, this is also the region where the rejection probabilities start overlapping (see Figure 4.5).

Additionally, Figure 4.13 validates our judgment about the rough granularity of the redundancy for $\frac{E_S}{N_0} > -6\ dB$ based on the $P(R_j)$ vs. $\frac{E_S}{N_0}$ plot in Section 4.2.4 (there we based the judgment on the nonoverlapping rejection probabilities). In the $\overline{n_{trans}}$ vs. $\frac{E_S}{N_0}$ plot, a system overdesign is reflected in the stair-like shape of the average transmission curve. A better transmission of redundancy granularity would have resulted in a smooth increase of the $\overline{n_{trans}}$ curve.

4.4 Throughput

The transmission rates of commercial communication products are usually given in information bits per time, as for example 56 $\frac{kbit}{sec}$ for a modem. For the error performance, however, the actual symbol duration is of no importance. Instead, the time independent signal energy to noise power density ratio $\frac{E_S}{N_0}$ is the crucial variable. The aim of this section is to define and analyze a time independent measure for the on average transmitted information bits per symbol: the throughput. The actual temporal data rate will be treated in Section 4.7.

4.4.1 Definition

In Section 4.1 the coding/modulation rate was defined as a measure of the amount of informations bits per modulation symbol of a FEC system. For these systems, this rate was a fixed number. We also already introduced the coding-/modulation rate sequence $R_{CM}^{ARQ,j}$, $j \geq 1$ of an ARQ system parameter, measuring the amount of redundancy transmitted after the j-th transmission.

However, the actual amount of symbols which are on average required by a given ARQ system to transmit an information packet of length L_{info} reliably over a channel with a certain condition, is dependent on the coding-/modulation rate sequence $R_{CM}^{ARQ,j}$, $j \geq 1$ and the distribution of the number of transmission for that channel condition. Hence, like the number of transmissions $\overline{\mathbf{n}_{trans}}$, the total number $\mathbf{L}_{tot}$ of channel symbols required for the transmission of one information block of length L_{info} is a random variable. The quotient between L_{info} and the average value of $\mathbf{L}_{tot}$ is the actual coding-/modulation rate at which the ARQ system works for a given channel condition. This value will be called the throughput T. More precisely:

> The throughput T of a digital communication system is defined as the quotient of an information block length L_{info} in *bit* and the expected value of channel symbols $E\{\mathbf{L}_{tot}\}$ in *symbol* which are required by the receiver to transmit this information block without errors:

$$T = \frac{L_{info}}{E\{\mathbf{L}_{tot}\}}. \tag{4.20}$$

The author likes to draw attention to the phrase 'which are required by the receiver' in the throughput definition. Accordingly, only the symbols, which were transmitted *and* used for decoding, are actual counted. As a result, the throughput is independent of the transmission process and purely a function of the coding and combining scheme incorporated into the ARQ system. This definition differs in that respect from some of the many throughput definitions found in literature. The author decided for this definition since it nicely splits the data rate performance of an ARQ system in two separate parts:

- the information theoretical performance (characterized by the throughput as defined above)
- the effects of the transmission scheme and environmental parameters such as roundtrip delays, symbol durations, etc.

From the statement 'to transmit ... without errors' follows that throughput can not be a property of a digital FEC communication system, since each FEC system has some amount of residual error rate. Nevertheless, each FEC has a certain coding-/modulation rate and a certain data rate.

For ARQ systems, successive retransmissions increase the amount of transmitted symbols but without changing the amount of transmitted information bits. If no retransmissions occur, the throughput has its maximum value T_{max}, which is equivalent to the coding/modulation rate of

the first packet

$$\begin{aligned} T_{Max} &= R_{CM}^{ARQ,1} \\ &= R_{CM}^{FEC,1}. \end{aligned}$$

As an example, a memoryless CE-ARQ system employing a rate $\frac{1}{3}$ FEC code and 16-ary QAM (4 bit/symbol) has a maximum throughput of

$$T_{max} = R_{CM}^{FEC} \frac{bit}{symbol} = \frac{1}{3} \cdot 4 \frac{bit}{symbol} = \frac{4}{3} \frac{bit}{symbol}.$$

As the number of on average required transmissions $\overline{n_{trans}}$ increase, the throughput T decreases and, due to the monotonically decreasing coding-/modulation rate, finally approaches it limit value zero. Section 4.4.4 provides a detailed analysis of the throughput.

Again, the author likes to emphasis that the throughput is measured on a basis of channel symbols, which were actually used for decoding. Possible transmission gaps as well as symbols transmitted without being used for the decoding process are not taken into account. Hence, the throughput is independent of the applied transmission processes (SW, GBN, and SR) as defined in Section 2.1. The advantage of this throughput definition is that it solely describes the coding theoretical performance of the coding and combining scheme employed in the ARQ system for which theoretical upper bounds (see Section 4.4.2) can be found. The applied transmission process may lead to an additional system performance degradation, which comes on top of the throughput performance. Section 4.7.3 discusses this relation of the throughput and the actual data rate.

4.4.2 Channel Capacity

In his famous 1948 paper, Claude E. Shannon [Sha48] introduced the channel capacity as maximum average mutual information

$$C = \max_{f_{\mathbf{x}}(x)} I(\mathbf{x};\mathbf{y}) \tag{4.21}$$

which can be transmitted over a noisy channel, where the maximization is performed over the distribution $f_{\mathbf{x}}(x)$ of the input symbols x.

Descriptive, the channel capacity C determines the maximum rate of information bits per channel symbol with which information can be transmitted over a noisy channel with arbitrary small error probability. On the other hand, if the rate of information bits per channel symbol is larger than the capacity C of the channel, it is by any means impossible to transmit error free. Shannon illustrated this in his original paper ([Sha48]) with the help of Figure 4.14, depicting an ideal correcting system. An observer notes the errors in the received information and transmits correcting data to the receive point. Now, if an attempt is made to transmit the information M with a rate larger than the channel capacity then *"Nature takes payment by requiring just as much uncertainty* (correcting data), *so that we are actually not getting more than C through correctly"* [Sha48].

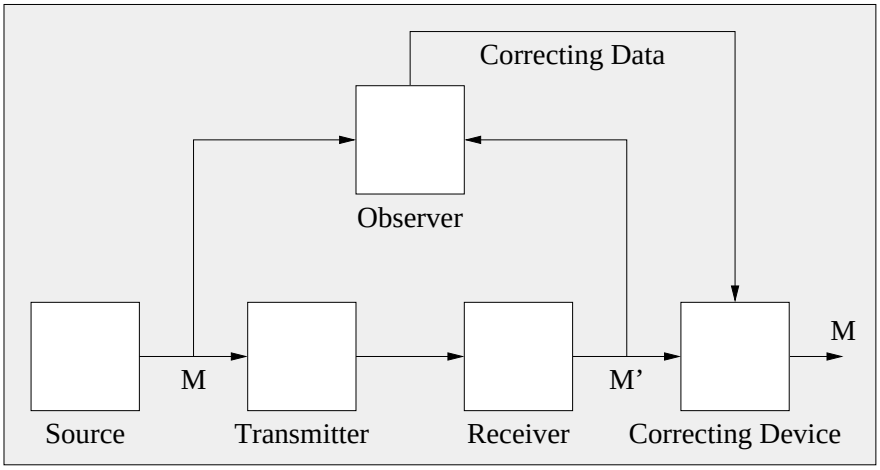

Figure 4.14: Correcting System. Taken from Shannon's 1948 Paper[Sha48].

As already mentioned in Section 2.3, in that sense Figure 4.14 depicts a perfect decision feedback system. Correct received information is delivered to the receive point and for erroneous data just as much additional information as required for its correction is requested. Hence, the channel capacity represents the maximum possible information bit to channel symbol rate, which can be transmitted error free by any digital communication system over a wireless channel with a certain quality. Therefore, the channel capacity C is the upper bound to the throughput T as defined in Section 4.4.1. The absence of the ideal observer and the limitation to decision feedback in an actual ARQ systems leads to additional performance degradations, since

- the exact locations of the errors are unknown
- the required amount of additional information is unknown.

For the one dimensional AWGN channel Equation 4.21 is maximized for a continuous input alphabet with normal distribution, yielding [Kre89]

$$C_{1D-AWGN}\left(\frac{E_S}{N_0}\right) = \frac{1}{2} \cdot log_2\left(1 + 2 \cdot \frac{E_S}{N_0}\right) \frac{bit}{symbol}. \tag{4.22}$$

Equivalently, the channel capacity for the two dimensional case is achieved for a two dimensional normal distributed input alphabet and given by [Kre89]

$$C_{2D-AWGN}\left(\frac{E_S}{N_0}\right) = log_2\left(1 + \frac{E_S}{N_0}\right) \frac{bit}{symbol}. \tag{4.23}$$

The channel capacities for the one and two-dimensional AWGN channel are shown in Figure 4.15.

As mentioned, these capacities are obtained under a set of assumptions, among them an infinite channel symbol alphabet. In practical digital communication systems, however, the allowed

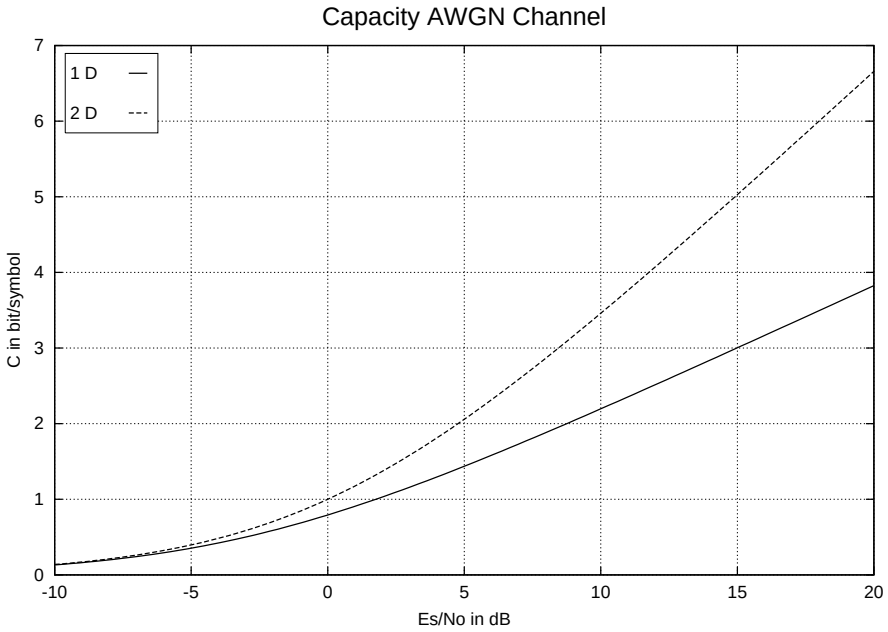

Figure 4.15: Channel Capacities for AWGN Channel

transmission signals x_j are restricted to a certain set of M symbols (M-ary modulation). Under these circumstances the in Figure 4.15 depicted capacities can not be achieved, even not theoretically (as an example, if a binary modulation format is used, 1 bit/symbol can be transmitted at maximum). If one regards the modulation as part of a new channel model and under the assumption that all symbols are equal probable the maximization over $f_{\mathbf{x}}(x)$ in Equation 4.21 is unnecessary and the new capacities of the channel with the incorporated modulation can be obtained as the average mutual information. In [Kre89] equations for the capacities of M-ary QAM and PSK ($M < 64$) over an AWGN channel are calculated and solved numerically.

4.4.3 $\frac{E_S}{N_0}$ versus $\frac{E_b}{N_0}$

Section 4.1 already dealt with the topic of $\frac{E_S}{N_0}$ versus $\frac{E_b}{N_0}$. However, so far we only considered system properties which are not a direct function of the throughput T, such as the BER of an FEC system or the rejection probabilities $P(R_j)$ of an ARQ system. For these system properties, the performance measure vs. $\frac{E_S}{N_0}$ plot differed on by a horizontal shift from the corresponding plot vs. $\frac{E_b}{N_0}$.

The throughput T of an ARQ system, on the other hand, is obviously a function of T. And also the channel capacity C - as maximum throughput R_{max} - is a function of R_{max}. Not only that for these quantities the plot versus $\frac{E_b}{N_0}$ is not simply a shift compared to the plot versus $\frac{E_S}{N_0}$, but it is not even given that such a plot exits. In the following we consider the AWGN channel capacities and the throughputs of an ARQ system.

Channel Capacity vs. $\frac{E_b}{N_0}$

The channel capacities of the AWGN channel are given by Equations 4.22 and 4.23 as a function of $\frac{E_S}{N_0}$. Substituting the channel capacity C for the coding-/modulation rate in Equation 4.5 yields an expression for $\frac{E_S}{N_0}$ in terms of the normalized signal-energy-to-noise-power-density ratio:

$$\frac{E_S}{N_0} = C \cdot \frac{E_b}{N_0}. \tag{4.24}$$

Using this equation to substitute $\frac{E_S}{N_0}$ in the expressions for the channel capacity yields a fixed point equation

$$\begin{aligned} C &= f\left(\tfrac{E_S}{N_0}\right) \\ &= f\left(C \cdot \tfrac{E_b}{N_0}\right) \end{aligned} \tag{4.25}$$

In Appendix C.4.1 it is shown that the 1-dimensional and the 2-dimensional AWGN channel Equation 4.25 has a unique fixed point for $\frac{E_b}{N_0} > ln(2)$ and $R_{max} = C > 0$. Hence, the channel capacities for the AWGN channel can be represented as a function of the normalized signal-energy-to-noise-power-density ratio $\frac{E_b}{N_0}$:

$$C = f\left(\frac{E_b}{N_0}\right). \tag{4.26}$$

$\frac{E_b}{N_0}_{min} = ln(2)$ represents the famous smallest $\frac{E_b}{N_0}$ which must be exceeded to enable reliable transmission (although only if $R \to 0$). In logarithmic measure

$$\left.\frac{E_b}{N_0}_{min}\right|_{dB} = 10 \cdot log_{10}\left(ln\left(2\right)\right)\ dB = -1.59\ dB.$$

The two channel capcities $C_{1D-AWGN}\left(\frac{E_b}{N_0}\right)$ and $C_{2D-AWGN}\left(\frac{E_b}{N_0}\right)$ are depicted in Figure 4.16.

Throughput vs. $\frac{E_b}{N_0}$

As with the channel capacity, if the argument $\frac{E_S}{N_0}$ of the throughput $T = f\left(\frac{E_S}{N_0}\right)$ of an ARQ system is replaced with $T \cdot \frac{E_b}{N_0}$ also a fixed point equation is obtained:

$$T = f\left(T \cdot \frac{E_b}{N_0}\right). \tag{4.27}$$

Contrary to the AWGN channel capacities, this fixed point equation has not necessarily a unique solution for any $R > 0$ and $\frac{E_b}{N_0} > \left.\frac{E_b}{N_0}\right|_{min}$. The reason why the AWGN channel capacities can be expressed as function of $\frac{E_S}{N_0}$ can be found in its steadily increasing slope (second derivative is positive for all $\frac{E_S}{N_0}$ values), resulting in one unique fixed point (see Appendix C.4.1). An arbitrary ARQ system, however, does not have a throughput with steadily increasing slope. Instead, typical throughput curves of ARQ systems have a stair-like shape with alternating decreasing and

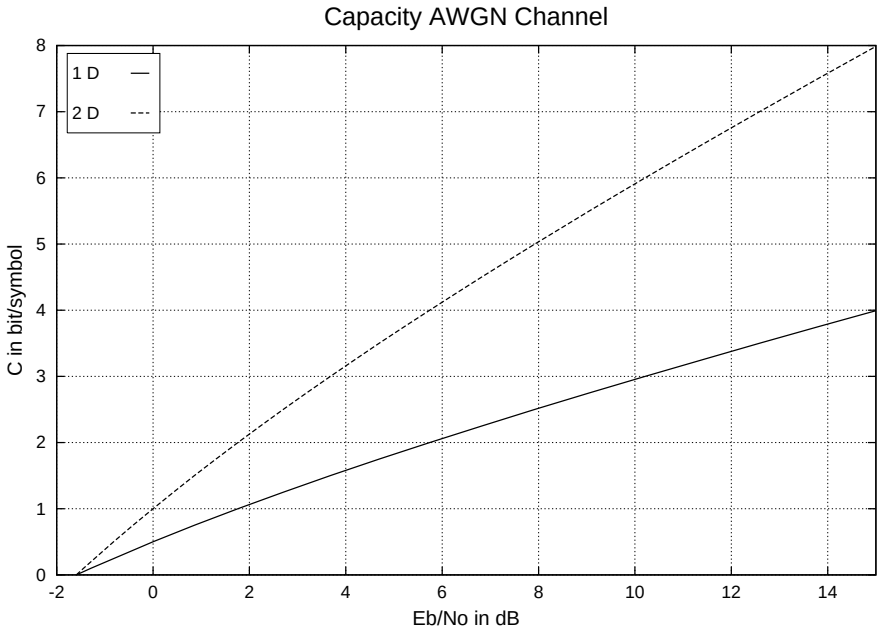

Figure 4.16: AWGN Channel Capacities versus $\frac{E_b}{N_0}$.

increasing slopes (second derivative takes positive and negative values). To illustrate the effect, Figure 4.17 shows the left side $y = x$ of the Fixed Point Equation 4.27 and also the right side $y = T\left(\frac{E_b}{N_0} \cdot x\right)$ for three different $\frac{E_b}{N_0}$ values. The throughput curve $T\left(\frac{E_S}{N_0}\right)$ is taken from simulations of a punctured rate-$\frac{1}{3}$ convolutional coded VE-MARQ system with information block length 280 bit and coding-/modulation rate sequence of $\frac{280}{288} \cdot \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{6}, \frac{1}{9}, \ldots\right\}$ (corresponding rejection probabilities in Figure 4.5, average number of transmissions in Figure 4.13, and throughput in Figure 4.21). Depending on the $\frac{E_b}{N_0}$, there are variable numbers of fixed points. As with the channel capacity, there is a minimum $\frac{E_b}{N_0}$ in order to have one fixed point ($\frac{E_b}{N_0} = 0.09\,dB$, for example, is smaller than this minimum value). If the normalized signal-to-noise is large enough there is, due to the constant throughput for large $\frac{E_S}{N_0}$, only one unique fixed point (for example at $\frac{E_b}{N_0} = 8.33\,dB$). But there are also some normalized signal-to-noise values with several fixed points (at least 3 fixed points for $\frac{E_b}{N_0} = 6.20\,dB$). This example already shows that a throughput of an arbitrary ARQ system can not be expressed as a **function** of $\frac{E_b}{N_0}$. Nevertheless, due to the steadyness of the throughput $y = T\left(\frac{E_b}{N_0} \cdot x\right)$ for all $\frac{E_b}{N_0} > 0$ and $x > 0$, we can obtain a steady curve showing the **relation** between T and $\frac{E_b}{N_0}$. Figure 4.18 depicts this relation for the VE-MARQ system. Each individual cross at a certain $\frac{E_b}{N_0}$ is one solution of the fixed point equation. As it can be already taken from Figure 4.17, there are 3 different rates with the same normalized signal-to-noise ratio $\frac{E_b}{N_0} = 6.20\,dB$. That means, that there are 3 operational points where the system uses up the same amount of energy per information bit, yet with different throughputs.

By looking at Figure 4.18 two major dents become apparent: one when the throughput drops from its maximum value to about 0.5 and another one from rates 0.32 to 0.16. These dents are indications of a bad system design: the rate drop due to the additional redundancy results in a larger

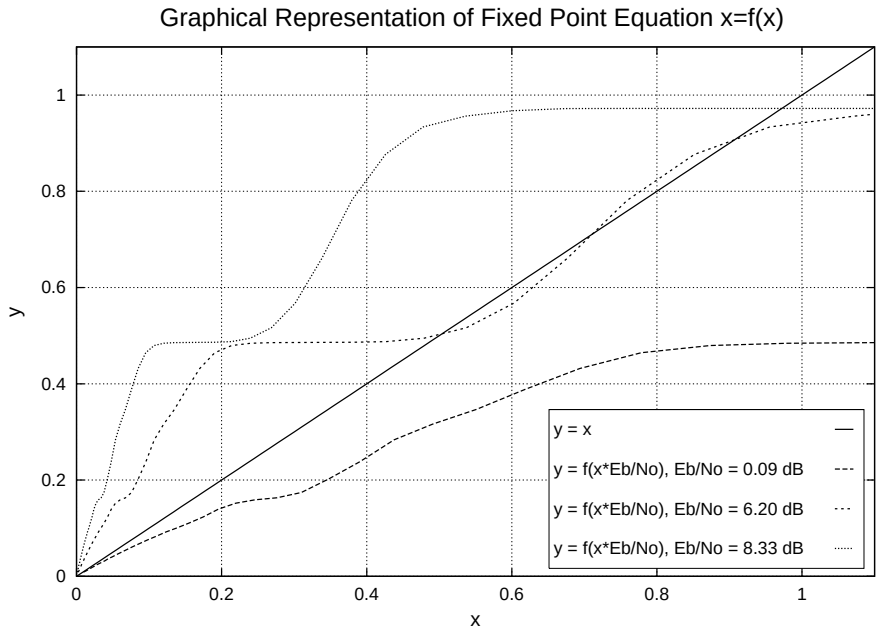

Figure 4.17: Illustration of Fixed Point Equation 4.27 for a Convolutional Coded ARQ System. See Figure 4.21 for Corresponding Throughput.

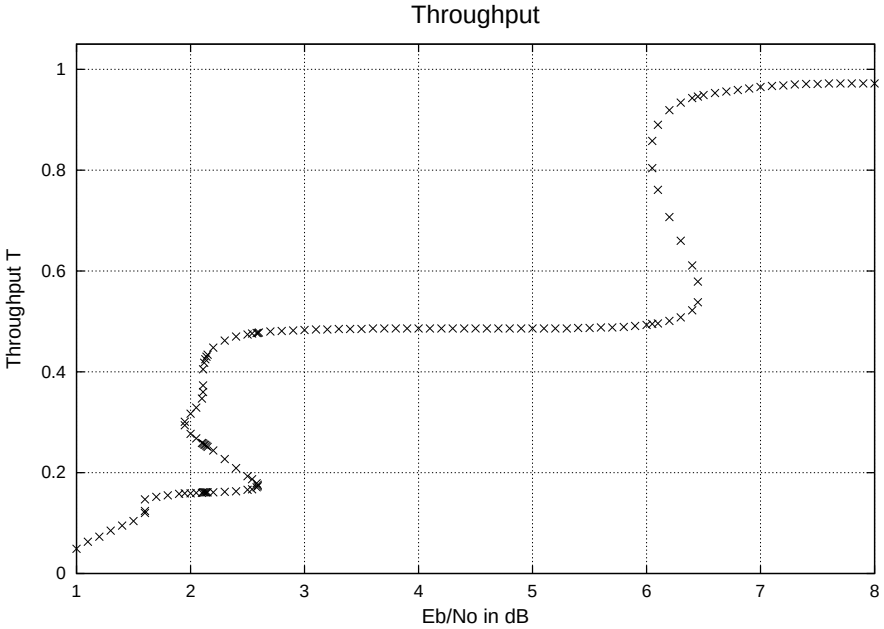

Figure 4.18: Relation Between T and $\frac{E_b}{N_0}$Resulting from the Solution of the Fixed Point Equation 4.27 for the HARQ System with Throughput Depicted in Figure 4.21.

shift towards higher $\frac{E_b}{N_0}$ values than the shift towards smaller $\frac{E_b}{N_0}$ values due to the increased error correction capability, i.e. the system has a negative coding gain for these rates. In general, these dents in a T vs. $\frac{E_b}{N_0}$ plot are an indication for a too coarse granularity of the delivered redundancy. In our example, the first dent results from the rate drop from $R_{CM}^{ARQ,1} = \frac{280}{288}$ to $R_{CM}^{ARQ,2} = \frac{280}{288} \cdot \frac{1}{2}$ and the second one from the rate drop from $R_{CM}^{ARQ,3} = \frac{280}{288} \cdot \frac{1}{3}$ to $R_{CM}^{ARQ,4} = \frac{280}{288} \cdot \frac{1}{6}$.

From all discussed indications for system overdesign ($P(R_j)$ vs. $\frac{E_S}{N_0}$, $\overline{n_{trans}}$ vs. $\frac{E_S}{N_0}$, and in this section T vs. $\frac{E_b}{N_0}$) the throughput versus the normalized signal-energy-to-noise-power-density ratio reveals most dramatically the inefficiency of ARQ systems for some operational points. A well designed system, which delivers redundancy in small increments has, like the channel capacity, a functional relation between the throughput and $\frac{E_b}{N_0}$.

Although the T vs. $\frac{E_b}{N_0}$ provides useful information theoretical insight into an ARQ system, the throughput vs. $\frac{E_S}{N_0}$ still provides sufficient information for comparison of ARQ schemes. These plots are easier to obtain and - contrary to the BER versus $\frac{E_S}{N_0}$ plot of a FEC system - takes all important system parameters into account and therefore provides a fair comparison: The compared systems use the same bandwidth (since the plot is based on $\frac{E_S}{N_0}$) and the different rates of the systems are also apparent (the throughput T itself).

4.4.4 Throughput Performance

Objective of this subsection is to relate the throughput T with other ARQ system performance measures defined in the previous sections. To do so, we need to make use of already introduced parameters, namely the length of the information packet to be transmitted L_{info} in bit and the lengths L_j of the packets sent during the j-th transmissions in $symbol$. If all transmissions use the same packet length, it will again simply be denoted as L. This particular case is easier to handle and will treated first before we move on to variable packet lengths.

Constant Packet Length L

If the packet length is constant and equal to L, the random variable total number of symbols $\mathbf{L}_{tot}$ can take only values $n \cdot L,\ n \geq 1$ with a non-vanishing probability. Hence,

$$P(\mathbf{L}_{tot} = x) = \begin{cases} P(\mathbf{n}_{trans} = n) & , x = n \cdot L \\ 0 & , otherwise \end{cases}$$

and

$$\begin{aligned} E\{\mathbf{L}_{tot}\} &= \sum_{n=1}^{\infty} n \cdot L \cdot P(\mathbf{L}_{tot} = n \cdot L) \\ &= L \cdot \sum_{n=1}^{\infty} n \cdot P(\mathbf{n}_{trans} = n) \\ &= L \cdot \overline{n_{trans}}\,. \end{aligned} \tag{4.28}$$

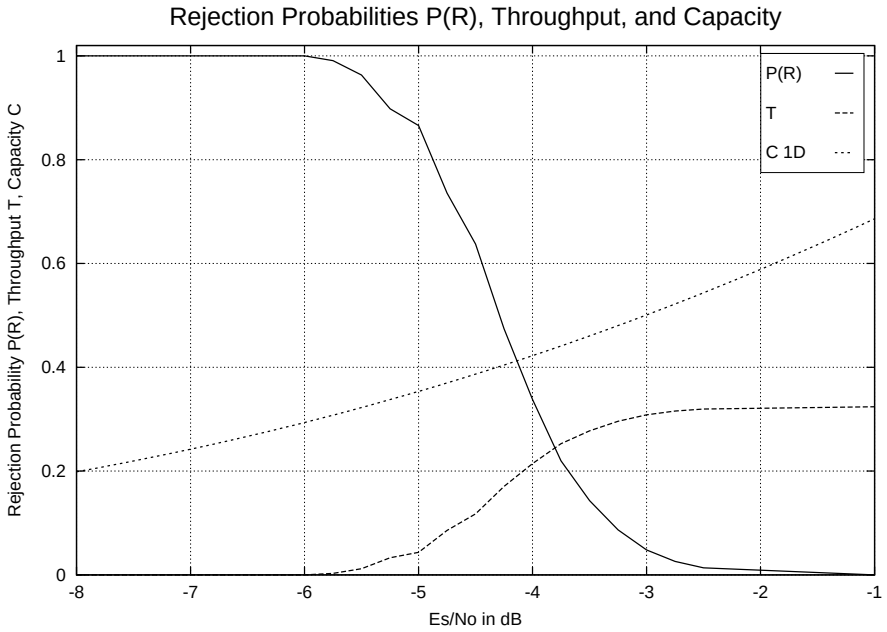

Figure 4.19: Rejection Probability and Throughput of a Memoryless CE-ARQ System with Coding-/Modulation Rate $R_{CM}^{ARQ,j} = \frac{280}{288} \cdot \frac{1}{3 \cdot j}$ Over AWGN.

Substituting Equation 4.28 into Equation 4.20 yields

$$T = \frac{L_{info}}{L \cdot \overline{n_{trans}}}$$

and since the packet size is constant, Equation 4.4 can be applied

$$T = \frac{R_{CM}^{FEC}}{\overline{n_{trans}}}. \tag{4.29}$$

Hence, the throughput of any ARQ system with constant packet sizes can immediately be derived by dividing the coding-/ modulation rate R_{CM}^{FEC} of the embedded FEC system with the average number of transmissions $\overline{n_{trans}}$.

For **memoryless CE-ARQ** systems follows from Equation 4.17 and Equation 4.29

$$T = R_{CM}^{FEC} \cdot (1 - P(R)). \tag{4.30}$$

During the discussion of the shapes of the rejection probability in Section 4.2.3, it was mentioned that $P(R)$ varies within a few dB from 0 to 1 as one moves towards lower $\frac{E_S}{N_0}$-values (see Figure 4.5). According to Equation 4.30, the throughput of this system falls within that region from its maximum thoughput R_{CM} to zero. This behavior and its relation to the rejection probability can be seen in Figure 4.19, which depicts the AWGN channel capacity and the throughput of an ARQ system with its corresponding rejection probability.

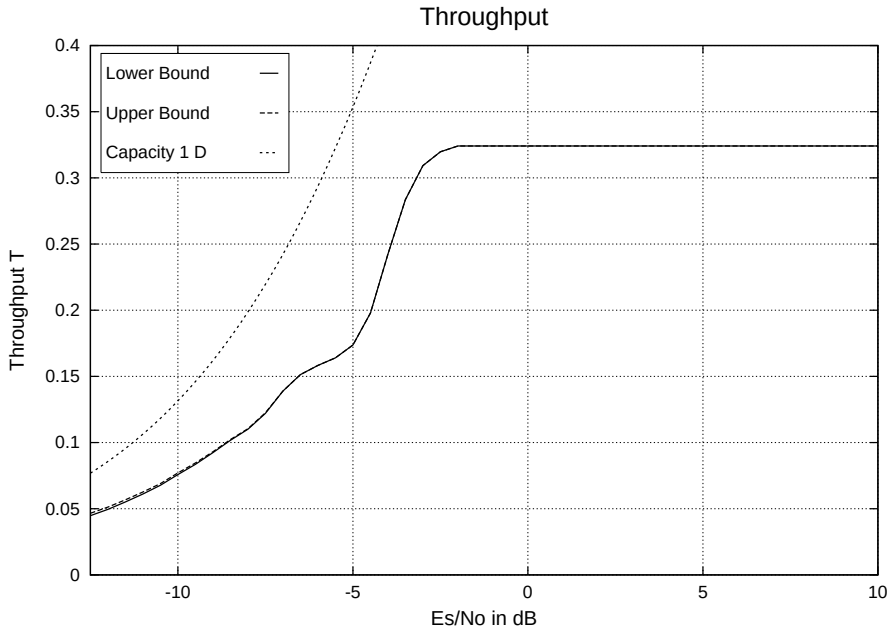

Figure 4.20: Bounds of the Throughput of a CE-MARQ System with Coding-/ Modulation Rate $R_{CM}^{ARQ,j} = \frac{280}{288} \cdot \frac{1}{3 \cdot j}$ over AWGN.

For **CE-MARQ** systems, Inequation 4.19 can be used to bound the throughput:

$$\frac{R_{CM}^{FEC}}{1 + \sum_{k=1}^{\infty} P(R_k)} \leq T \leq \frac{R_{CM}^{FEC}}{1 + \sum_{k=1}^{\infty} \left[\prod_{i=1}^{k} P(R_i) \right]}$$

Figure 4.20 depicts the the bounds on the throughput of a CE-MARQ system in conjunction with the channel capacity. The tightness of the bounds for the average number of transmissions results also in extreme tight bounds for the throughput.

Figures 4.19 and 4.20 show the throughput of two identical ARQ systems. The only difference is that the decoder of the second one makes use of all previous transmission, i.e. these two system differ only in the receiver realization. It can be seen, that the MARQ realization outperforms the memoryless version as soon as a third transmission becomes likely. Whereas the memoryless version has a vanishing throughput at $\frac{E_S}{N_0} = -6\,dB$, the MARQ realization has still a throughput of 50 % of its maximum value.

Variable Packet Lengths L_j

If the packet during the j-th transmission consists of L_j symbols, the random variable $\mathbf{L}_{tot}$ can take only values $\sum_{j=1}^{n} L_j,\ n \geq 1$ with non-vanishing probability. Hence,

$$P(\mathbf{L}_{tot} = x) = \begin{cases} P(\mathbf{n}_{trans} = n) & , x = \sum_{j=1}^{n} L_j \\ 0 & , otherwise \end{cases} \tag{4.31}$$

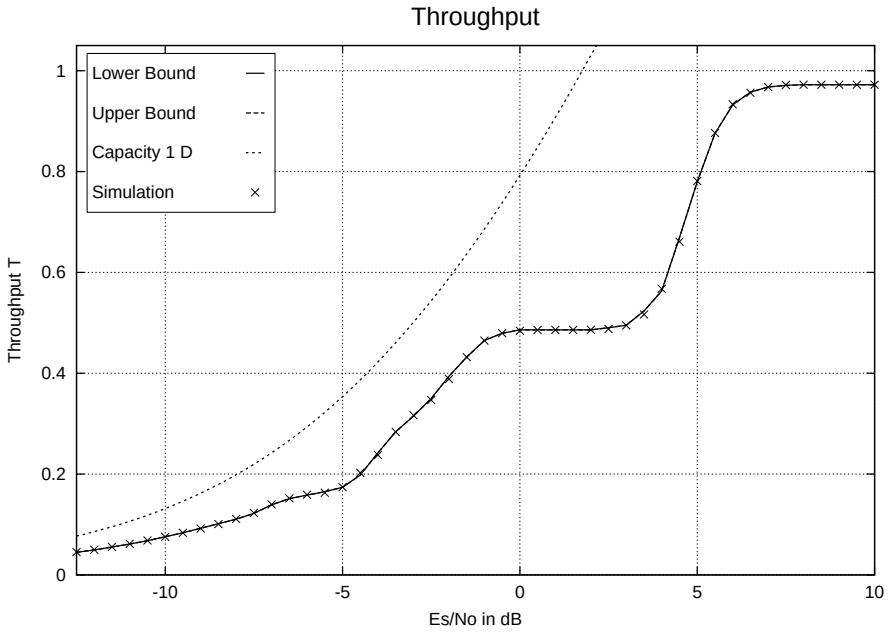

Figure 4.21: Simulation and Bounds of the Throughput of a VE-MARQ System with Coding-/ Modulation Rates $R_{CM}^{ARQ,j} = \frac{280}{288} \cdot \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{6}, \ldots, \frac{1}{3 \cdot (j-2)}, \ldots\right\}$ over AWGN.

In Appendix C.4.2 it is shown

$$\begin{aligned} E\{\mathbf{L}_{tot}\} &= L_1 + L_2 \cdot P(R_1) + L_3 \cdot P(R_1 \cap R_2) + \cdots \\ &= L_1 + \textstyle\sum_{n=1}^{\infty} L_{n+1} \cdot P\left(\bigcap_{i=1}^{n} R_i\right) \end{aligned} \tag{4.32}$$

Using Equation 4.32 and the bounds for the probability $P\left(\bigcap_{i=1}^{n} R_i\right)$ (see Inequation 4.9),

$$\prod_{i=1}^{n} P(R_i) \leq P\left(\bigcap_{i=1}^{n} R_i\right) \leq P(R_n),$$

bounds for the expected value of transmitted symbols can be derived:

$$L_1 + \sum_{n=1}^{\infty} \left[L_{n+1} \cdot \prod_{i=1}^{n} P(R_i) \right] \leq E\{\mathbf{L}_{tot}\} \leq L_1 + \sum_{n=1}^{\infty} \left[L_{n+1} \cdot P(R_n) \right]. \tag{4.33}$$

Substituting Inequation 4.33 in Equation 4.20 finally yields bounds for the throughput T of an ARQ system with unequal packet lengths:

$$\frac{L_{info}}{L_1 + \sum_{n=1}^{\infty} \left[L_{n+1} \cdot P(R_n) \right]} \leq T \leq \frac{L_{info}}{L_1 + \sum_{n=1}^{\infty} \left[L_{n+1} \cdot \prod_{i=1}^{n} P(R_i) \right]} \tag{4.34}$$

or

$$\frac{1}{\frac{1}{R_{CM}^{FEC,1}} + \sum_{n=1}^{\infty}\left[\frac{1}{R_{CM}^{FEC,n+1}} \cdot P\left(R_n\right)\right]} \leq T \leq \frac{1}{\frac{1}{R_{CM}^{FEC,1}} + \sum_{n=1}^{\infty}\left[\frac{1}{R_{CM}^{FEC,n+1}} \cdot \prod_{i=1}^{n} P\left(R_i\right)\right]}.$$

Contrary to the ARQ systems with constant packet sizes, there is no functional relation of the throughput with the average number of transmissions. Figure 4.21 shows the simulated throughput of a VE-MARQ system with uneven packet sizes (due to varying puncture rates) together with the corresponding upper and lower bounds as well as the channel capacity. Again, the bounds are extremely close, which arises from the hardly overlapping rejection probabilities, but also of the inverse character of Equation 4.34.

4.5 Delay

In an ARQ system, several delays exist. In the following two important delays are examined in more detail: the average packet delay and the average information delay. So far in this chapter, no temporal properties of the system were of concern. Now, however, the transmission protocols introduced in Section 2.1 and several timings occurring in the system play a decisive role. A crucial time is the round trip delay t_{RT}. The round trip delay summarizes all delays occurring from the transmission of the *last* packet symbol at the transmitter until the ACK/NAK is received and has been processed by the transmitter. Specifically, the round trip delay comprises the two propagation delays from the transmitter to the receiver, the forward propagation $t_{prop}^{TX,RX}$ and the backward propagation time $t_{prop}^{RX,TX}$, the duration of an ACK or NAK t_{ACK} , and finally both processing times at receiver t_{prc}^{RX} and transmitter t_{prc}^{TX}:

$$t_{RT} = t_{prop}^{TX,RX} + t_{prc}^{TX} + t_{ACK} + t_{prop}^{RX,TX} + t_{prc}^{TX}.$$

Obviously some of this times may vary throughout the retransmission process, yet the round trip delay t_{RT} will be treated as a constant for the remainder of this work.

Besides the round trip delay itself, its ratio $n_{RT}^{(j)}$ with the individual packet durations $t_p^{(j)}$ is also an important characteristics: In Section 4.4.4 the length L_j of the j-th packet in $symbol$ was already defined. With the constant symbol duration t_S, it follows for the duration of the individual packets

$$t_p^{(j)} = L_j \cdot t_S \tag{4.35}$$

and for the mentioned ratio

$$n_{RT}^{(j)} = \frac{t_{RT}}{t_p^{(j)}} = \frac{t_{RT}}{t_S \cdot L_j}. \tag{4.36}$$

The ratio $n_{RT}^{(j)}$ indicates the number of packets of type j, which fit physically on the round trip route. Let n_{RT}^{min} and n_{RT}^{max} denote the minimum and maximum of $n_{RT}^{(j)}$ for future derivations.

Systems with constant packet sizes, on the other hand, have only one constant ratio, denoted

$$n_{RT} = \frac{t_{RT}}{t_S \cdot L}.$$

Even for ARQ systems with varying channel packet sizes, the information packet size L_{info} is a constant. In analogy with the above definitions, we define the so-called round trip number N_{RT} as the quotient of the round trip delay and the duration of an information packet:

$$N_{RT} = \frac{t_{RT}}{t_S \cdot L_{info}}. \tag{4.37}$$

This value is an actual environment parameter and independent of the ARQ type and its realization and also independent of the ARQ protocol. It will be seen that the round trip number plays an important role for the various delays and the actual data rate of an ARQ system . In general the relation

$$n_{RT}^{(k)} = R_{CM}^{FEC,k} \cdot N_{RT}$$

holds and more specific for systems with constant packet sizes

$$n_{RT} = R_{CM}^{FEC} \cdot N_{RT}. \tag{4.38}$$

In order to get a feeling of the order of N_{RT} for typical wireless systems we consider in the following section two exemplary environments.

4.5.1 Examples of Wireless Transmission Systems

In the following, two typical wireless transmission links, which considerably differ concerning the round trip delay, are presented. As one extreme, a link via a geostationary satellite with high data rate is regarded. On the other hand, the second example is a low range line-of-sight (LoS) link with as of today average data rate. Both systems will be used through the remainder of this section to illustrate the derived results.

Satellite Link

For a high data rate satellite link, we consider Digital Video Broadcasting (DVB) over a geostationary satellite (distance from Frankfurt, Germany $\approx$ 37600 km). If we neglect the ACK/NAK duration and the processing times, the total round trip delay is 4 times the propagation time from the transmitter to the satellite, hence,

$$t_{RT} = \frac{4 \cdot h}{c} = \frac{4 \cdot 37600\,km}{3 \cdot 10^8\,\frac{m}{sec}} = 0.5\,sec$$

The highest symbol rate, specified in the DVB standard [DVB97] is $42.2 \cdot 10^6$ QPSK symbols per second. Accordingly, the specified constant information packet length of $204\ byte$ results in a

packet length of

$$L = \frac{204 \cdot 8\,bit}{2 \cdot \frac{bit}{symbol}} = 816\,symbol$$

and a packet duration of

$$t_p = \frac{816\,symbol}{42.2 \cdot 10^6 \cdot \frac{symbol}{sec}} = 19.3\,\mu sec$$

Consequently, the number of packets on the round trip route is

$$n_{RT} = \frac{0.5\,sec}{19.3\,\mu sec} = 25900,$$

which means that 25900 packets fit on the round trip route. The round trip number is obtained to

$$\begin{aligned} N_{RT} &= \frac{n_{RT}}{R_{CM}^{FEC}} \\ &= \frac{25900}{2} \\ &= 12950\,. \end{aligned}$$

This example ranks due to its high data rate among the systems with the highest round trip numbers. In general the round trip number for satellite links starts with 10 for very low rate connections and extends up to 15. For the remainder of this work a sample satellite link with a comparable small round trip number of $N_{RT} = 30$ and $n_{RT} = 30$ is chosen.

Line-of-Sight Link

Now, we consider a typical line-of-sight link with a distance of $4\,km$ between transmitter and receiver, a information data rate of $1\,\frac{Mbit}{sec}$, and a typical ARQ information block length of $200\,bit$. Hence, the above parameters are obtained to

$$t_{RT} = \frac{2 \cdot d}{c} = \frac{2 \cdot 4\,km}{3 \cdot 10^8\,\frac{m}{sec}} = 26.67\,\mu sec,$$

$$t_p = \frac{200\,bit}{1 \cdot 10^6 \cdot \frac{bit}{sec}} = 0.2\,msec,$$

and

$$n_{RT} = \frac{26.67\,\mu sec}{0.2\,msec} = 0.13$$

and with the assumption of a coding-/modulation rate of 1 it also follows for the round trip number

$$N_{RT} = 0.13.$$

Contrary to the above satellite system, not even one packet fits onto the round trip route. In

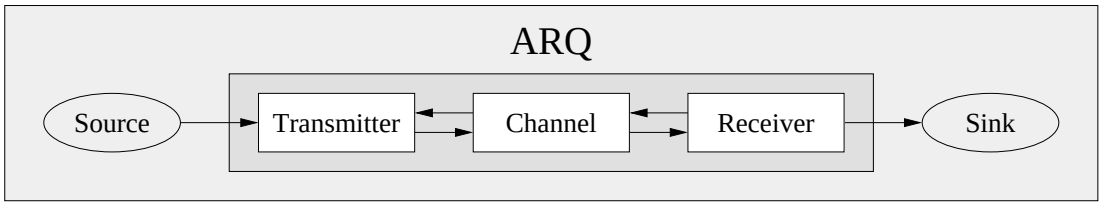

Figure 4.22: Schematic Block Diagram of the Source - ARQ - Sink Link.

general for terrestrial digital communication systems, the round trip number is by orders smaller than for satellite links, since data rates and packet sizes are comparable, whereas the distance is by orders higher. Typically the round trip number ranges

$$N_{RT} = 10^{-4} \ldots 10.$$

For the remainder of this work an exemplary line-of-sight link with a round trip number $N_{RT} = 0.2$ and $n_{RT} = 0.2$ is chosen.

4.5.2 Average Packet Delay

The ARQ system with the channel provides the link between the information source and the sink (see Figure 4.22). It takes information packets from the source and releases them to the sink after the error check has been passed. After a packet is transfered from the source to the ARQ system it needs to be stored for possible retransmission, whereas the source can delete that particular packet. After a certain time, the acknowledgment for that packet is received by the transmitter and also the transmitter of the ARQ system can erase that packet. The following definition of the average packet delay $\overline{\Delta t_p}$ aims for this time:

> The random variable packet delay $\Delta\mathbf{t}_p$ of a certain packet is the time from the instant of the first transmission of the first symbol until - after possible retransmissions - finally the ACK for that packet is received by the transmitter. Its expected value will be called the average packet delay $\overline{\Delta t_p}$.

In the following, bounds for $\overline{\Delta t_p}$ in dependency of the rejection probabilities are derived for the three protocols defined in Section 2.1.

Stop-and-Wait Protocol

Feature of the SW protocol is that after the first transmission of a packet the transmitter remains idle until the acknowledgment is received. This means that for the first transmission the time $t_1 + t_{RT}$ elapses. In case of a second transmission, the time $t_1 + t_2 + 2 \cdot t_{RT}$ passed by until the second ACK/NAK is received. In the n-th transmission the time increment $t_p^{(n)} + t_{RT}$ is added. Therefore, the packet delay $\Delta\mathbf{t}_p$ can take only values $\sum_{j=1}^{n} t_p^{(j)} + n \cdot t_{RT},\ n \geq 1$ with

the corresponding probabilities $P(\mathbf{n}_{trans} = n)$. Hence, the probability distribution of the random variable $\Delta\mathbf{t}_p$ for a SW protocol is

$$P(\Delta\mathbf{t}_p = x) = \begin{cases} P(\mathbf{n}_{trans} = n) & , x = \sum_{j=1}^{n} t_p^{(j)} + n \cdot t_{RT} \\ 0 & , otherwise. \end{cases} \tag{4.39}$$

In Appendix C.5.1 it is shown that the resulting average packet delay $\overline{\Delta t_p}$ is bounded by

$$t_{RT} + L_1 \cdot t_S + \sum_{n=1}^{\infty} \left[(t_{RT} + L_{n+1} \cdot t_S) \cdot \prod_{i=1}^{n} P(R_i) \right] \tag{4.40}$$

$$\leq \quad \overline{\Delta t_p}\big|_{SW} \tag{4.41}$$

$$\leq \quad t_{RT} + L_1 \cdot t_S + \sum_{n=1}^{\infty} \left[(t_{RT} + L_{n+1} \cdot t_S) \cdot P(R_n) \right] \tag{4.42}$$

and that the normalized average packet delay $\frac{\overline{\Delta t_p}}{t_{RT}}$ is related to the throughput and the average number of transmissions by

$$\left.\frac{\overline{\Delta t_p}}{t_{RT}}\right|_{SW} = \overline{n_{trans}} + \frac{1}{N_{RT}} \cdot \frac{1}{T}. \tag{4.43}$$

The first addend of Equation 4.43 is the delay resulting from the multiple roundtrips, whereas the second addend origins from the time it takes to transmit the individual packets. The environment parameter N_{RT} brings the two contributions into relation and heavily determines which one of the two parts in Equation 4.43 dominates. This becomes already clear if we determine the minimal average packet delay for perfect channel conditions. Then,

$$\overline{n_{trans}} \quad = \quad 1$$

$$T \quad = \quad R_{CM}^{ARQ,1} = R_{CM}^{FEC,1}$$

and therefore

$$\left.\frac{\overline{\Delta t_p}}{t_{RT}}\right|_{min} \quad = \quad 1 + \frac{1}{N_{RT} \cdot R_{CM}^{FEC,1}}.$$

For systems with constant packet sizes of L *symbol* the Bounds 4.40 - 4.42 simplify to

$$(t_{RT} + L \cdot t_S) \cdot \left[1 + \sum_{k=1}^{\infty} \left[\prod_{i=1}^{k} P(R_i) \right] \right] \leq \overline{\Delta t_p}\big|_{SW} \leq (t_{RT} + L \cdot t_S) \cdot \left[1 + \sum_{k=1}^{\infty} P(R_k) \right].$$

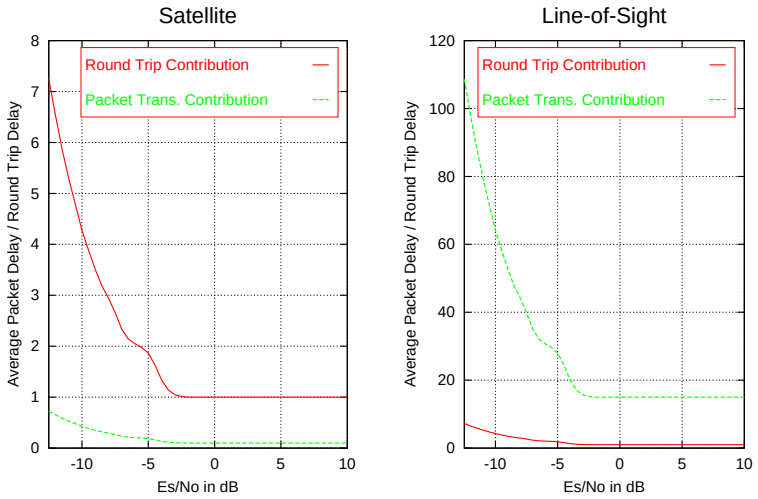

Figure 4.23: Normalized Average Packet Delays of the CE-MARQ System with Throughput Depicted in Figure 4.20 and SW Protocol.

Substituting Equation 4.29 into Equation 4.43 yields for CE-ARQ systems

$$\begin{aligned} \overline{\frac{\Delta t_p}{t_{RT}}}\Big|_{SW} &= \overline{n_{trans}} + \frac{1}{N_{RT}} \cdot \overline{\frac{n_{trans}}{R_{CM}^{FEC}}} \\ &= \left(1 + \frac{1}{n_{RT}}\right) \cdot \overline{n_{trans}}\,. \end{aligned} \tag{4.44}$$

Equation 4.44 more clearly reflects the two separate components of the total average packet delay: the round trip contribution ($t_{RT} \cdot \overline{n_{trans}}$) and the contribution from the packet durations

$$\frac{t_{RT}}{n_{RT}} \cdot \overline{n_{trans}} = t_p \cdot \overline{n_{trans}}.$$

When $n_{RT} = 1$, that is that the packet duration is identical with the round trip delay, both parts contribute to equal parts to the average packet delay. However, for practical systems this in most cases not even approximately the case. Instead, we have the following two cases:

- $n_{RT} \ll 1$, i.e the packet duration is much longer than the round trip delay, mainly the packet duration part dominates.
- $n_{RT} \gg 1$, i.e. the round trip delay is much longer than the packet duration, the round trip contribution is the dominating part.

These two obvious behaviors are reflected in Equation 4.44. As a conclusion, for the majority of terrestrial systems the average packet delay will be due to the time it takes to transmit the packets and for the majority of satellite systems due to the round trip delay.

Figure 4.23 depicts both addends of Equation 4.43 for the two example wireless systems defined in Section 4.5.1. As expected, the round trip delay contribution dominates in the satellite system, whereas the packet duration plays the important role in the line-of-sight system.

As the throughput T, the average packet delay $\overline{\Delta t_p}$ is a function of the average number of transmission $\overline{n_{trans}}$ for systems with constant packet sizes. Consequently, $\overline{\Delta t_p}$ can also be expressed as a function of the throughput. Substituting Equation 4.29 into Equation 4.44 yields

$$\begin{aligned} \frac{\overline{\Delta t_p}}{t_{RT}} &= \left(1 + \frac{1}{n_{RT}}\right) \cdot \frac{R_{CM}^{FEC}}{T} \\ &= \left(R_{CM}^{FEC} + \frac{1}{N_{RT}}\right) \cdot \frac{1}{T} \,. \end{aligned}$$

Go-Back-N Protocol

In the above discussed SW protocol, the average packet delay was constant for all information packets transfered from the source to the ARQ system. As we will see, this is in general not the case for the GBN protocol. Therefore, let $\overline{\Delta t_p}^{(k)}$ depict the average packet delay of the k-th packet entering the ARQ system.

The varying average packet delay arises from the fact, that the transmitter of an ARQ system with GBN protocol sends packets on a permanent basis. To illustrate this, consider packet #2. The packet is immediately sent after its only predecessor, the first packet. However, the first decoding attempt of packet #2 is made only after the first packet is successfully decoded, i.e. it is possibly transmitted several times needlessly. This time adds to the packet delay since the packet was already transfered to the ARQ system. The same argumentation applies to packet #3, but this time the two preceding packets need to be accepted by the receiver. Since that probability is smaller, the additional time is on average larger than the corresponding time for the packet #2. However, if a decoding attempt is made with an arbitrary packet is not dependent on all previously sent packet but rather only on the packets whose acknowledgments have not been received. As a conclusion we expect the average packet delay of a GBN protocol ARQ to increase monotonously up to the packet $\left\lfloor n_{RT}^{(1)} \right\rfloor + 1$. Thereafter it remains constant.

In Appendix C.5.1 the following expression for the average packet delay $\overline{\Delta t_p}^{(k)}$ of the k-th packet transfered to the ARQ is derived:

$$\left.\frac{\overline{\Delta t_p}^{(k)}}{t_{RT}}\right|_{GBN} = \begin{cases} k \cdot \overline{n_{trans}} - k + 1 + k \cdot \dfrac{1}{N_{RT}} \cdot \dfrac{1}{T} - \dfrac{k-1}{n_{RT}^{(1)}} & , \; k \leq \left\lceil n_{RT}^{(1)} \right\rceil + 1 \\ \left(\left\lceil n_{RT}^{(1)} \right\rceil + 1\right) \cdot \overline{n_{trans}} - \left\lceil n_{RT}^{(1)} \right\rceil & \\ \quad + \dfrac{\left\lceil n_{RT}^{(1)} \right\rceil + 1}{N_{RT}} \cdot \dfrac{1}{T} - \dfrac{1}{R_{CM}^{FEC,1}} & , \; k > \left\lceil n_{RT}^{(1)} \right\rceil + 1 \end{cases} \tag{4.45}$$

Equation 4.45 indeed reflects the increase of the average packet delay up to packet $\left\lceil n_{RT}^{(1)} \right\rceil + 1$.

The first packet has always an average packet delay, which is identical to that of the SW protocol:

$$\frac{\overline{\Delta t_p}^{(1)}}{t_{RT}} = \overline{n_{trans}} + \frac{1}{N_{RT}} \cdot \frac{1}{T}.$$

The second packet has always an increased average packet delay of

$$\frac{\overline{\Delta t_p}^{(2)}}{t_{RT}} = 2 \cdot \overline{n_{trans}} + 1 + 2 \cdot \frac{1}{N_{RT}} \cdot \frac{1}{T} - \frac{1}{n_{RT}^{(1)}}.$$

Whether the following packets have a further increased average packet delay depends on the round trip number N_{RT}. If no more than one packet fits on the round trip route ($N_{RT} = 1$), all further average packet delays are identical to $\overline{\Delta t_p}^{(2)}$.

In order to bound the average round trip delay $\overline{\Delta t_p}^{(k)}$ with the help of the rejection probabilities, the appropriate bounds of $\overline{n_{trans}}$ (Inequations 4.19 on Page 70) and T (Inequations 4.34 on Page 82) are applied to Equation 4.44 yielding for $k \leq \left\lceil n_{RT}^{(1)} \right\rceil + 1$

$$\begin{aligned} & 1 + \frac{1}{n_{RT}^{(1)}} + k \cdot \sum_{n=1}^{\infty} \left[\left(1 + \frac{1}{n_{RT}^{(n+1)}} \right) \cdot \prod_{i=1}^{n} P\left(R_i\right) \right] \\ & \leq \left. \frac{\overline{\Delta t_p}^{(k)}}{t_{RT}} \right|_{GBN} \leq \\ & 1 + \frac{1}{n_{RT}^{(1)}} + k \cdot \sum_{n=1}^{\infty} \left[\left(1 + \frac{1}{n_{RT}^{(n+1)}} \right) \cdot P\left(R_n\right) \right] \end{aligned} \tag{4.46}$$

and for $k > \left\lceil n_{RT}^{(1)} \right\rceil + 1$

$$\begin{aligned} & 1 + \frac{1}{n_{RT}^{(1)}} + \left(\left\lceil n_{RT}^{(1)} \right\rceil + 1 \right) \cdot \sum_{n=1}^{\infty} \left[\left(1 + \frac{1}{n_{RT}^{(n+1)}} \right) \cdot \prod_{i=1}^{n} P\left(R_i\right) \right] \\ & \leq \left. \frac{\overline{\Delta t_p}^{(k)}}{t_{RT}} \right|_{GBN} \leq \\ & 1 + \frac{1}{n_{RT}^{(1)}} + \left(\left\lceil n_{RT}^{(1)} \right\rceil + 1 \right) \cdot \sum_{n=1}^{\infty} \left[\left(1 + \frac{1}{n_{RT}^{(n+1)}} \right) \cdot P\left(R_n\right) \right]. \end{aligned} \tag{4.47}$$

Again, for constant packet durations $t_p = t_S \cdot L$ Equation 4.45 and Bounds 4.46 and 4.47 simplify to

$$\left. \frac{\overline{\Delta t_p}^{(k)}}{t_{RT}} \right|_{GBN} = \begin{cases} \left(k \cdot \overline{n_{trans}} - k + 1 \right) \cdot \left(1 + \dfrac{1}{n_{RT}} \right) & , \; 0 < k \leq \lceil n_{RT} \rceil + 1 \\ \left(\left(\lceil n_{RT} \rceil + 1 \right) \cdot \overline{n_{trans}} - \lceil n_{RT} \rceil \right) \cdot \left(1 + \frac{1}{n_{RT}} \right) & , \; k > \lceil n_{RT} \rceil + 1 \end{cases}$$

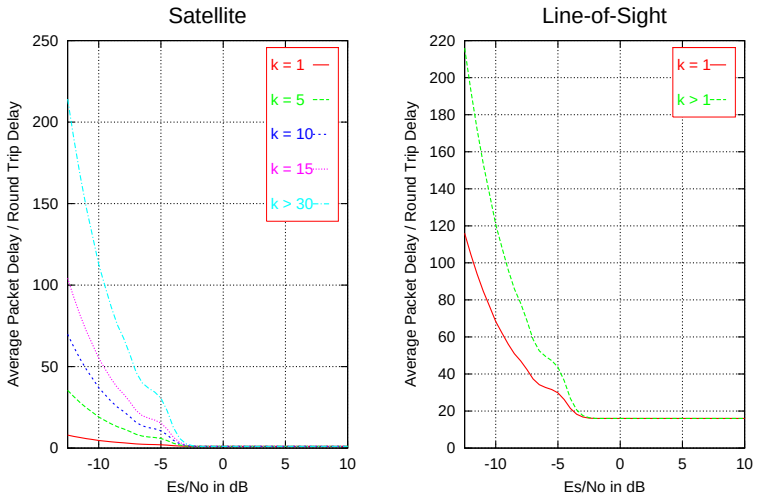

Figure 4.24: Normalized Average Packet Delay in Dependency of the Packet Number k of the CE-MARQ Systems with Throughput Depicted in Figure 4.20 and GBN Protocol.

and for $0 < k \leq \lceil n_{RT} \rceil + 1$

$$\left(1 + k \cdot \sum_{j=1}^{\infty} \left[\prod_{i=1}^{j} P\left(R_i\right)\right]\right) \cdot \left(1 + \frac{1}{n_{RT}}\right)$$

$$\leq \left.\frac{\overline{\Delta t_p}^{(k)}}{t_{RT}}\right|_{GBN} \leq$$

$$\left(1 + k \cdot \sum_{j=1}^{\infty} P\left(R_j\right)\right) \cdot \left(1 + \frac{1}{n_{RT}}\right)$$

and for $k > \lceil n_{RT} \rceil + 1$

$$\left(1 + (\lceil n_{RT} \rceil + 1) \cdot \sum_{j=1}^{\infty} \left[\prod_{i=1}^{j} P\left(R_i\right)\right]\right) \cdot \left(1 + \frac{1}{n_{RT}}\right)$$

$$\leq \left.\frac{\overline{\Delta t_p}^{(k)}}{t_{RT}}\right|_{GBN} \leq$$

$$\left(1 + (\lceil n_{RT} \rceil + 1) \cdot \sum_{j=1}^{\infty} P\left(R_j\right)\right) \cdot \left(1 + \frac{1}{n_{RT}}\right)$$

This two average packet delays are depicted in Figure 4.24 for the satellite and the line-of-sight scheme presented in Section 4.5.1.

The same figure also depicts some of the average packet delays of the discussed satellite system. If this figure and Figure 4.23 are compared, it becomes apparent that the GBN protocol for

environments with a large round trip delay (e.g. satellite systems) has a multiple of the packet delay compared to the SW protocol. This is the price one has to pay (packets need to be stored for that time at the transmitter) for possible performance gains which will be discussed later.

Selective-Repeat Protocol

In the last section, we saw that the GBN protocol has an increased average packet delay compared to the SW protocol. This arose from the fact, that the ARQ system with GBN protocol accepted packets from the source and transmitted them, however occasionally without performing a decoding attempt. If the SR protocol is used, on the other hand, a decoding attempt is made with all sent packets. Therefore, for an individual packet there is no difference between the SW and the SR protocol. The only difference is that the ARQ system with SR uses the time for transmission of additional packets, where the SW protocol has transmission gaps. From that, one can conclude that the average packet delay of the SR and the SW protocol are identical. Hence,

$$\left.\frac{\overline{\Delta t_p}}{t_{RT}}\right|_{SR} = \overline{n_{trans}} + \frac{1}{n_{RT}} \cdot \frac{1}{T}$$

and

$$1 + \frac{1}{n_{RT}^{(1)}} + \sum_{n=1}^{\infty}\left[\left(1 + \frac{1}{n_{RT}^{(n+1)}}\right) \cdot \prod_{i=1}^{n} P\left(R_i\right)\right] \leq \left.\frac{\overline{\Delta t_p}}{t_{RT}}\right|_{SR} \leq 1 + \frac{1}{n_{RT}^{(1)}} + \sum_{n=1}^{\infty}\left[\left(1 + \frac{1}{n_{RT}^{(n+1)}}\right) \cdot P\left(R_n\right)\right]$$

$$1 + \frac{1}{n_{RT}^{(1)}} + \sum_{n=1}^{\infty}\left[\left(1 + \frac{1}{n_{RT}^{(n+1)}}\right) \cdot \prod_{i=1}^{n} P\left(R_i\right)\right] \leq \left.\frac{\overline{\Delta t_p}}{t_{RT}}\right|_{SR} \leq 1 + \frac{1}{n_{RT}^{(1)}} + \sum_{n=1}^{\infty}\left[\left(1 + \frac{1}{n_{RT}^{(n+1)}}\right) \cdot P\left(R_n\right)\right]$$

$$1 + \frac{1}{n_{RT}^{(1)}} + \sum_{n=1}^{\infty}\left[\left(1 + \frac{1}{n_{RT}^{(n+1)}}\right) \cdot \prod_{i=1}^{n} P\left(R_i\right)\right] \leq \left.\frac{\overline{\Delta t_p}}{t_{RT}}\right|_{SR} \leq 1 + \frac{1}{n_{RT}^{(1)}} + \sum_{n=1}^{\infty}\left[\left(1 + \frac{1}{n_{RT}^{(n+1)}}\right) \cdot P\left(R_n\right)\right].$$

4.5.3 Average Information Delay

In the previous section, the average delay of a single packet was considered. This delay is an important time for the system designer for the discussed reasons. However, for the ARQ user it is of less concern. She or he is more interested in the delay of the information frames to be transmitted with the ARQ system. We therefore define L_I as the length of the information frame in bit and

$$N_I = \left\lceil \frac{L_I}{L_{info}} \right\rceil$$

as the number of packets per information frame. We define the information frame delay $\Delta\mathbf{t}_I$ as:

> The random variable information frame delay $\Delta\mathbf{t}_I^{(N_I)}$ of a certain information frame of size N_I times the packet size is defined as the time from the instant of the first transmission of the first symbol until the ACK of the last packet is received. Its expected value will be called the average information delay $\overline{\Delta t}_I^{(N_I)}$.

In the following $\overline{\Delta t}_I^{(N_I)}$ will be investigated for the three ARQ protocols.

Stop-and-Wait Protocol

The SW protocol transmits the individual packets one after the other, i.e. packet n is only transmitted if packet $n-1$ has been acknowledged. Therefore the random variable $\Delta\mathbf{t}_I^{(N_I)}$ is the sum of N_I independent packet delays $\Delta\mathbf{t}_p^{(j)},\ j = 1, \ldots, N_I$:

$$\Delta\mathbf{t}_I^{(N_I)} = \sum_{j=1}^{N_I} \Delta\mathbf{t}_p^{(j)}.$$

If we normalize the average information delay to the round trip delay t_{RT}, it follows

$$\begin{aligned}
\frac{\overline{\Delta t}_I^{(N_I)}}{t_{RT}} &= \frac{E\left\{\Delta\mathbf{t}_I^{(N_I)}\right\}}{t_{RT}} \\
&= \sum_{j=1}^{N_I} \frac{E\left\{\Delta\mathbf{t}_p^{(j)}\right\}}{t_{RT}} \\
&= N_I \cdot \frac{\overline{\Delta t_p}}{t_{RT}}
\end{aligned}$$

since all the expected values are identical. Substituting Equation 4.43 finally yields

$$\left.\frac{\overline{\Delta t}_I^{(N_I)}}{t_{RT}}\right|_{SW} = N_I \cdot \left(\overline{n_{trans}} + \frac{1}{N_{RT}} \cdot \frac{1}{T}\right). \tag{4.48}$$

Equation 4.48 states that if a Stop-and-Wait protocol is used, the average information frame delay is the number of packets per frame times the average packet delay. We expect the more sophisticated protocols GBN and SR to have a smaller average information delay.

Go-Back-N Protocol

Contrary to the SW protocol, the transmitter of the GBN protocol does not remain idle. Although not all of the transmitted packets are actually used for a decoding attempt, we do expect a gain of this protocol over to the SW protocol.

In Appendix C.5.2 it is shown that the normalized average information delay of an information frame consisting of N_I packets is given by

$$\left.\frac{\overline{\Delta t_I}^{(N_I)}}{t_{RT}}\right|_{GBN} = N_I \cdot \overline{n_{trans}} - N_I + 1 + N_I \cdot \frac{1}{N_{RT}} \cdot \frac{1}{T} \tag{4.49}$$

Equation 4.49 indeed reflects the smaller information delay of the GBN protocol compared to the SW protocol, although - as we saw - the GBN protocol has a higher average packet delay:

$$\left.\frac{\overline{\Delta t_I}^{(N_I)}}{t_{RT}}\right|_{GBN} = \left.\frac{\overline{\Delta t_I}^{(N_I)}}{t_{RT}}\right|_{SW} - N_I + 1 \tag{4.50}$$

In order to find an interpretation for Equation 4.49 we consider the cases when the GBN protocol is more efficient: If packet n was successfully decoded, the decoder tries to decode the immediately following packet $n+1$. If this attempt was also successful the GBN protocol saved exact one round trip delay in comparison to the SW protocol. This case can happen for $N_I - 1$ packets (the first packet has no predecessor). With on average $\overline{n_{trans}}$ transmissions per packet the described case happens on average in $N_I - 1$ cases.

According to Equation 4.50, the average information delays of GBN and SW have the same contribution from the packet delays and differ only by $(N_I - 1) \cdot t_{RT}$. From that one can conclude two things:

1. The GBN protocol provides only an observable gain if the round-trip contribution is more dominant than the packet duration contribution, e.g. for satellite systems
2. The GBN protocol provides only an observable gain if the information frame length is a large multiple of the information length, i.e. N_I is large.

Figure 4.25 depicts the comparison of the average information frame delay for SW and GBN for the in Section 4.5.1 discussed systems with different numbers of packets per frame.

Selective Repeat Protocol

Both two previously derived information frame delays consisted of two parts. One part origins from the packet durations and the second part from the round trip delays. The packet duration

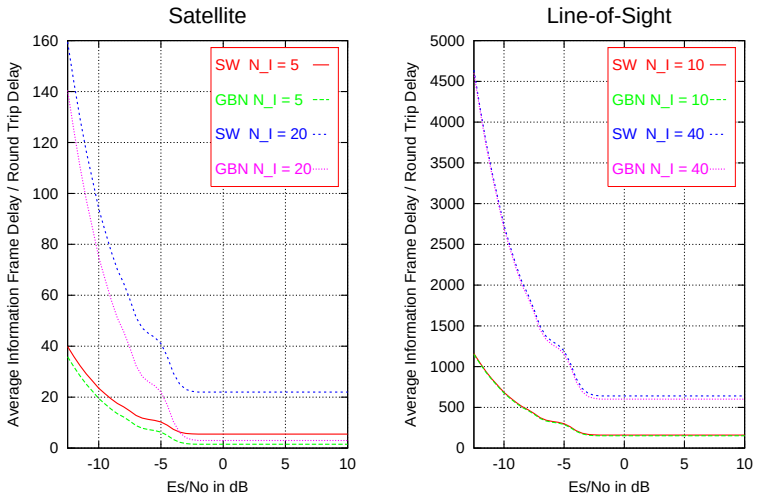

Figure 4.25: Comparison of the Normalized Average Information Frame Delay of the SW and GBN Protocol for Different Numbers of Packets per Frame.

part is the total time it takes to transmit all packets. This time

$$\left.\frac{\overline{\Delta t_I^{(N_I)}}}{t_{RT}}\right|_{packet\ duration} = \frac{N_I}{N_{RT}} \cdot \frac{1}{T}$$

is independent of the used transmission protocol. The round trip contribution, however, is transmission protocol dependent. It reflects the times when the transmitter is idle (for SW protocol) or transmits packets, which are not used for the decoding process (GBN protocol). With the SR protocol, where the transmitter continuously sends packets and all transmitted packets are used for the decoding process, one is tempted to conclude that there is no round trip delay contribution to the information frame delay for this protocol type. Yet for finite lengths of L_I a boundary effect occurs, which is illustrated in Figure 4.26. The first transmission gap occurs as soon as the tail of unacknowledged information packets has a length of the round trip number N_{RT} (time step 6). Depending on the following sequence of ACKs and NAKs several additional transmission gaps occur. The distribution of the number of transmission gaps is extremely difficult to derive. However, it seems obvious that the average number of transmissions $\overline{n_{trans}}$ and possibly the round trip number N_{RT} play an important role. The number of packets per information frame N_I, however, is of no concern for this protocol type. This is the basis for the performance gain of this protocol type for large information packets in terms of the information frame delay compared to the two other protocols. In Appendix C.5.2 the following bounds for the average information frame delay is reasoned[6]:

[6] The author likes to emphasis, that the upper bounds are not exactly derived, but rather heuristically obtained. Nevertheless, it has been verified for a variety of systems with different round trip numbers and different packets per information frame.

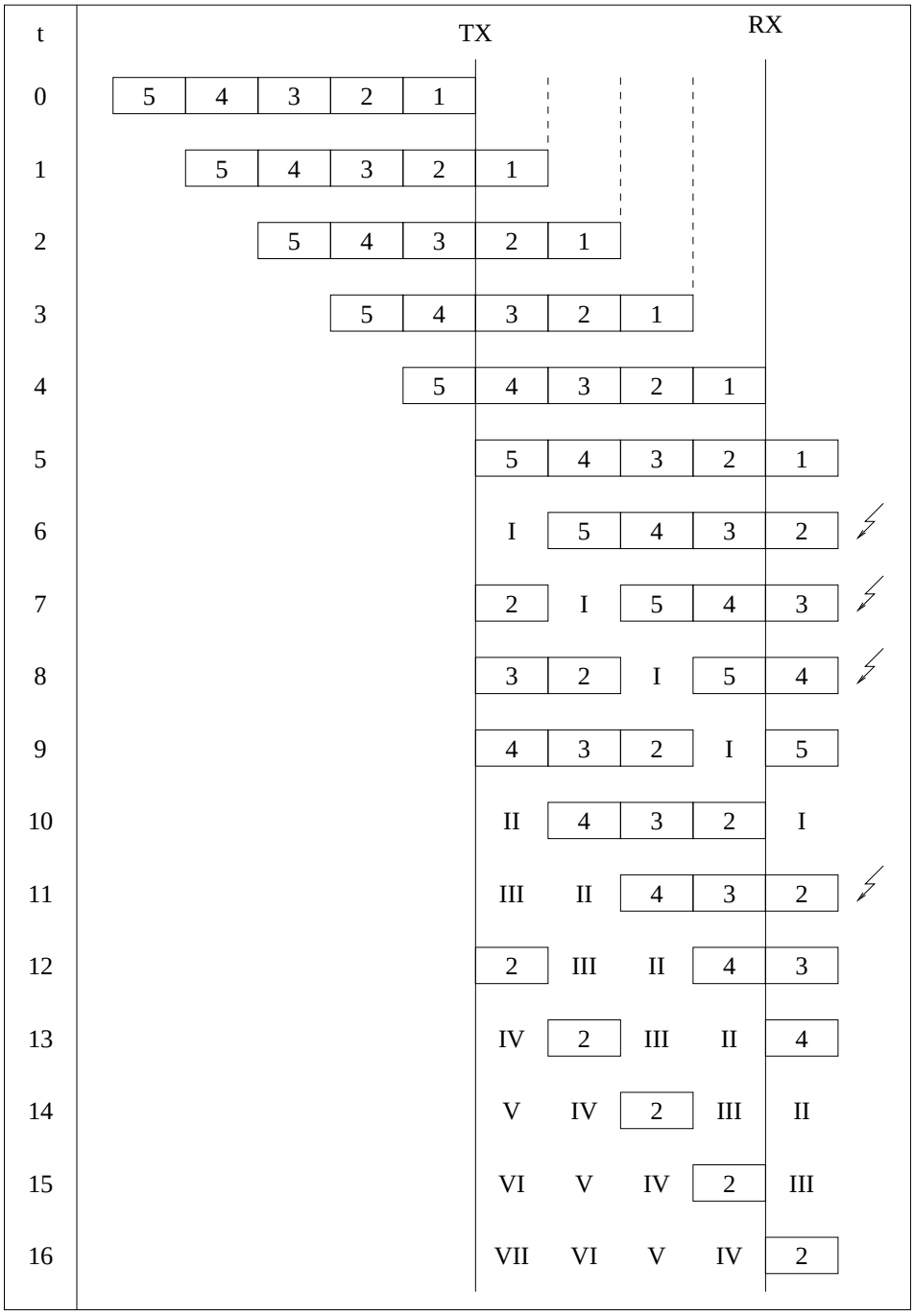

Figure 4.26: Illustration of the Boundary Effect in Transmission of Finite Packets with Selective Repeat Protocol.

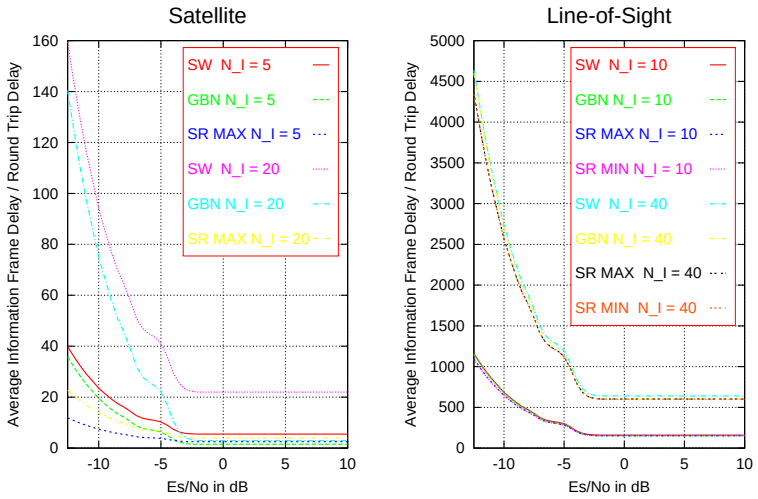

Figure 4.27: Comparison of the Normalized Average Information Frame Delay of the SW, GBN, and SR Protocol for Different Numbers of Packets per Frame.

$$1 + \frac{N_I}{N_{RT}} \cdot \frac{1}{T} \leq \left. \frac{\overline{\Delta t}_I^{(N_I)}}{t_{RT}} \right|_{SR} \leq 1 + \overline{n_{trans}} + \frac{N_I}{N_{RT}} \cdot \frac{1}{T} \tag{4.51}$$

By looking at Equation 4.51 it becomes apparent that the round trip contribution is indeed independent of the number of packets per information frame. Also, the number of packets on the round trip route N_{RT} plays a minor role and does accordingly not appear in the round trip delay contribution. However, as with the two previous protocols, $n_{RT}^{(1)}$ in Equation 4.51 determines which of the two contribution dominates. Figure 4.27 shows the performance of the three protocols for the two example systems of Section 4.5.1. As expected, the performance of all three protocols is almost identical for the terrestrial system, whereas the more sophisticated systems show performance gains in the satellite environment, especially for large information packet sizes.

4.6 Memory Requirements

The ARQ principle bases on retransmission for packets, which were not successfully decoded. In order to fullfill this task, the ARQ transmitter has to store the information to be transmitted until it is positively acknowledged. These transmitter memory requirements are investigated in the following section. In addition, also the receiver requires memory. These receiver memory requirements are analyzed in Section 4.6.2.

4.6.1 Transmitter Memory Requirements

It is the task of the transmitter to send the new packets as well as the retransmissions in case of negative acknowledgments. As soon as the ARQ system has accepted a packet from the source (see Figure 4.22), the transmitter is required to store this packet for possible retransmissions. More generally, the transmitter may store any form of information of a certain packet from which all possible retransmissions can be extracted.

The straight forward approach would be to store the information packet itself. Therefore a memory capable of storing L_{info} bits would be required. The advantage of this approach is that the required transmitter memory ranks among the smallest possible. On the other hand, the required computational load is among the largest, since for every retransmission, the information needs to be encoded and modulated again.

As another approach, the already endoded information sequence could be stored. Then, only the modulation is required to be performed several times for the retransmissions. However, the ARQ system is then restrained to working with that particular code and the freedom of choice in the retransmission method is limited. In addition, compared to storing the information sequence, more memory is required.

Software radios additionally provide the possibility of saving the already modulated baseband sequence. This is definitely the most memory extensive approach and also to most restrictive one concerning the choice of retransmissions. Only CE-ARQ implementations or VE-ARQ with a puncturing on channel symbol basis are possible. However, the computational load is also minimal, since encoding as well as base band modulation is required only once.

No matter what form of information is stored at the transmitter, each information packet requires a certain memory M_p. The maximum total transmitter memory requirement M_{tot}^{Tx} is consequently a multiple of this packet memory requirement. How many packets need to be stored at maximum is a function of the transmission protocol and the environmental and system parameters. As argued above, the individual packet memory requirement M_p^{Tx} is highly dependent on the transmitter realization. Hence, the normalized maximum transmitter memory requirement

$$\frac{M_{tot}^{Tx}}{M_p^{Tx}}$$

is introduced. In the following, this value is investigated for the various transmission protocols.

Stop-and-Wait Protocol

As previously discussed, the Stop-and-Wait protocol only accepts new information packet from the source if the current packet has been acknowledged. Hence, at any time only information for one packet needs to be stored an the normalized maximum transmitter memory requirement is

$$\left.\frac{M_{tot}^{Tx}}{M_p^{Tx}}\right|_{SW} = 1.$$

Due to the SW protocol, the transmitter memory requirements are independent system and environmental parameters.

Go-Back-N Protocol

The GBN protocol, in contrast to the SW protocol, continuously sends packets and has therefore to store information for all unacknowledged packets. In Section 4.5, the number of packets $\left\lceil n_{RT}^{(j)} \right\rceil$ of type j on the round trip route has already been defined. After the start of the transmission process, the ARQ system with GBN protocol continuously sends $\left\lceil n_{RT}^{(1)} \right\rceil$ packets before the first ACK or NAK is received. In case of a ACK, the first packet can be deleted from the transmitter memory and a be replaced with a new packet from the source. In case of a NAK, all packets need to be retransmitted: for the unacknowledged packet the first retransmission of length L_2 is sent, whereas for the remaining packets again the first transmission of size L_1 is used. If again the first packet is negatively acknowledged, the packet sequence is preceded by a packet of length L_3 and followed by packets of length L_1, and so on for further NAKs of packet one. If L_{min} denotes the minimal packet length and j_{min} the corresponding index, i.e.

$$L_{min} = \min_{j \geq 1} \{L_j\} = L_{j_{min}},$$

the maximum number of packets fits on the round trip if the preceding packet has this minimal length L_{min}. Then the remaining time for the following packets is $t_{RT} - t_S \cdot L_{min}$ and

$$\left\lceil \frac{t_{RT} - t_S \cdot L_{min}}{t_S \cdot L_1} \right\rceil$$

packets of size L_1 fit at least partly on the round trip route. Hence,

$$\left\lceil \frac{t_{RT} - t_S \cdot L_{min}}{t_S \cdot L_1} \right\rceil + 1$$

packets fit on the round trip route. In addition to the packets on the round trip route, one packet needs to be stored to provide continuous transmission and the required normalized maximum transmitter memory is

$$\left. \frac{M_{tot}^{Tx}}{M_p^{Tx}} \right|_{GBN} = \left\lceil \frac{t_{RT} - t_S \cdot L_{min}}{t_S \cdot L_1} \right\rceil + 2. \tag{4.52}$$

Usually the effect varying channel packet sizes can be neglected, especially for systems with a large round trip delay. For systems with constant channel packet sizes this effect is anyway not present and Equation 4.52 can be simplified to

$$\left. \frac{M_{tot}^{Tx}}{M_p^{Tx}} \right|_{GBN} = \lceil n_{RT} \rceil + 1.$$

Coming back to our sample satellite environments ($n_{RT} = 30$), the transmitter memory for

the satellite system would be 31, but for higher data rate systems it can easily exceed 1000. For the sample LoS environment where not even one channel packet fits on the round trip route, a maximum normalized transmitter memory of 2 is required.

Selective-Repeat Protocol

The difference between the SR and the GBN protocol is that the former individually retransmits erroneously detected packets. As a consequence, unlike the GBN protocol, the packet sequence on the round trip route can be composed of an arbitrary mixture of packets with different lengths. When all packets have the minimum length L_{min}, the maximum number $N_{RT}^{(j_{min})}$ of packets fits on the channel. Hence, the transmitter memory requirement for a SR protocol is

$$\left.\frac{M_{tot}^{Tx}}{M_p^{Tx}}\right|_{SR} = \left\lceil n_{RT}^{(j_{min})} \right\rceil + 1.$$

For constant packet sizes, the GBN and SR protocols have identical memory requirements of:

$$\left.\frac{M_{tot}^{Tx}}{M_p^{Tx}}\right|_{SR} = \lceil n_{RT} \rceil + 1.$$

4.6.2 Receiver Memory Requirements

In previous section the transmitter memory requirements were treated and found to be independent of operational system parameters such as the average number of transmissions or the troughput. This, however, is not in general the case for the receiver memory requirements. The reason for that can be found in the retransmissions. The receiver might have to store all or at least some of the already received information for one packet to aid the decoding process. Hence, the total receiver memory requirements are in this case a function of the required number of transmissions and the number of packets on the round trip route.

Obviously, the memory requirements at the receiver are heavily dependent on the implemented ARQ scheme as well on the actual ARQ receiver implementation, specifically what form of combining is used. Memoryless ARQ systems, on the other hand, have a constant receiver memory requirement per packet on the round trip route, since they discard all previously received information. Also CE-MARQ systems which combine each newly received channel packet with all previous channel packets to a new packet of constant size have also a constant receiver memory requirement per information packet. In general, however, the required memory for each information packet is dependent on the number of transmissions. As a description of this effect, we introduce the receiver memory cost function $M_p^{Rx}(k)$ and its normalized version

$$\frac{M_p^{Rx}(k)}{M_p^{Rx}}.$$

The receiver cost function $M_p^{Rx}(k)$ is the *overall* memory requirement for the information related

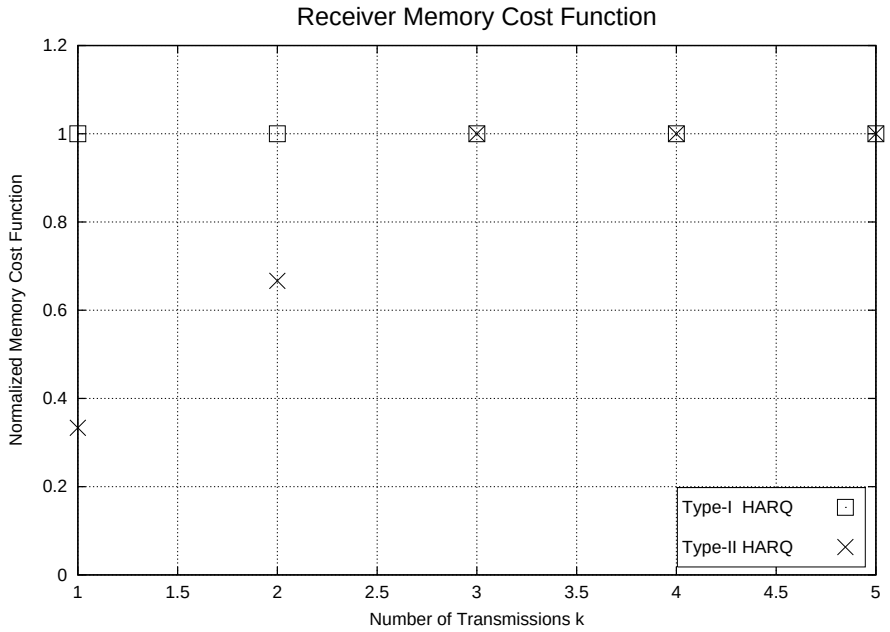

Figure 4.28: Normalized Receiver Memory Cost Function of a CE-MARQ and a VE-MARQ Scheme.

with *one* information packet at the receiver after the reception of the k-th transmission. The normalized version is a dimensionless quantity based on some normalization value M_p^{Rx}.

Figure 4.28 depicts the normalized receiver memory requirement of a CE-MARQ and the corresponding VE-MARQ scheme, obtained via puncturing. For the CE-MARQ scheme, a constant receiver memory equivalent to the normalization value M_p^{RX} is required, whereas the VE-MARQ scheme sends for the first three transmissions punctured packets. After the reception of the first packet only $\frac{1}{3} \cdot M_p^{Rx}$ is needed and after the reception of the second packet $\frac{2}{3} \cdot M_p^{Rx}$ (packet one is not a copy of packet one). Finally, after three transmissions, the mother code is achieved and all further transmissions are combined on a symbol basis, requiring the receiver memory M_p^{Rx} .

In Chapter 5 several methods for combining are introduced. For each scheme the memory cost function will be given. For the remainder of this section, all evaluations will be based on the general cost function $M_p^{Rx}(k)$.

Stop-and-Wait Protocol

With the SW protocol, only one packet is transmitted at a time. Hence, the receiver only needs to store the transmitted information associated with this information packet.

As a consequence, the total required receiver memory is identical to the receiver memory cost function. Systems with a constant memory cost function have therefore a normalized maximal

total receiver memory requirement of

$$\left.\frac{M_{tot}^{Rx}}{M_p^{Rx}}\right|_{SW} = 1. \tag{4.53}$$

On the other hand, if the memory cost function is increasing with the number of transmissions for one packet, the maximum required receiver memory is $M_p^{Rx}\left(n_{trans}^{max}\right)$ if n_{trans}^{max} denotes the maximum number of required transmissions. Since the number of required transmission $\mathbf{n}_{trans}$ is a random variable with the probability and cumulative distribution as discussed in Section 4.3.1, there is no maximum number of transmissions. Using the results from that section, however, we can determine an upper bound $n_{trans}^{max,upp}$ and a lower bound $n_{trans}^{max,low}$ for the maximal required transmissions with a certain probability P or error probability $P_E = 1 - P$. Then, the maximal required receiver memory M_{RX}^{max} under that constraint can be bounded as

$$\frac{M_p^{Rx}\left(n_{trans}^{max,low}\right)}{M_p^{Rx}} \leq \left.\frac{M_{tot}^{Rx}}{M_p^{Rx}}\right|_{SW} \leq \frac{M_p^{Rx}\left(n_{trans}^{max,upp}\right)}{M_p^{Rx}}. \tag{4.54}$$

For almost all practical ARQ systems, however, the memory cost function is not increasing infinitely. If $M_{p,max}^{Rx}$ denotes this maximum value, then an upper bound for the normalized total receiver memory can be derived:

$$\left.\frac{M_{tot}^{Rx}}{M_p^{Rx}}\right|_{SW} \leq \frac{M_{p,max}^{Rx}}{M_p^{Rx}}. \tag{4.55}$$

Go-Back-N Protocol

Although the SW and the GBN protocols considerably differ in the way how the transmitter send packets, both protocol types have in common that only one packet is decoded at a time and hence only the information for one packet needs to be kept in the receiver memory. As a conclusion, both protocol types have the identical receiver memory requirements and Equations 4.53 to 4.55 are also valid for ARQ systems with the GBN protocol.

Selective-Repeat Protocol

The two previously discussed protocol types have the same receiver memory requirements which arose from the fact that both protocol types keep only the transmitted information of one information packet in memory. The SR protocol, on the other hand, processes all received packets.

If the ARQ system uses **constant packet sizes**, $\lceil n_{RT} \rceil$ packets fit on the round trip route and hence it is that number for which receiver memory must be provided. If the system has constant receiver memory cost function, the total receiver memory is

$$\left.\frac{M_{tot}^{Rx}}{M_p^{Rx}}\right|_{SR} = \lceil n_{RT} \rceil .$$

For increasing memory cost functions, again the results of Section 4.3.1 can be used to bound the

normalized total receiver memory requirement

$$\frac{M_p^{Rx}\left(n_{trans}^{max,low}\right)}{M_p^{Rx}} \cdot \lceil n_{RT} \rceil \leq \left.\frac{M_{tot}^{Rx}}{M_p^{Rx}}\right|_{SR} \leq \frac{M_p^{Rx}\left(n_{trans}^{max,upp}\right)}{M_p^{Rx}} \cdot \lceil n_{RT} \rceil$$

and for limited memory cost functions $M_p^{Rx}(k) \leq M_{p,max}^{Rx}$ a general upper bound is

$$\left.\frac{M_{tot}^{Rx}}{M_p^{Rx}}\right|_{SR} \leq \frac{M_{p,max}^{Rx}}{M_p^{Rx}} \cdot \lceil n_{RT} \rceil .$$

For an ARQ system with **varying packet sizes**, two parameters are varying when retransmissions become likely. At first, the amount of packets which fit onto the round trip route is dependent on the consistency of the packets, i.e. how many packets of a certain type are in the queue and secondly, the required receiver memory for the individual packets (described by the receiver memory cost function). As mentioned, the memory cost function is monotonically increasing for practical systems, but $\left\lceil n_{RT}^{(j)} \right\rceil$ is totally dependent on the ARQ system. Hence the largest total receiver memory is required when the round trip route is totally composed of packets with maximum product

$$\left\lceil n_{RT}^{(j)} \right\rceil \cdot M_p^{Rx}(j) .$$

Hence, an equivalent bound for the normalized total receiver memory can be obtained to

$$\frac{\max_{j \leq n_{trans}^{max,low}} \left\{ \left\lceil n_{RT}^{(j)} \right\rceil \cdot M_p^{Rx}(j) \right\}}{M_p^{Rx}} \leq \left.\frac{M_{tot}^{Rx}}{M_p^{Rx}}\right|_{SR} \leq \frac{\max_{j \leq n_{trans}^{max,upp}} \left\{ \left\lceil n_{RT}^{(j)} \right\rceil \cdot M_p^{Rx}(j) \right\}}{M_p^{Rx}}.$$

4.7 Data Rate

In Chapter 4.4 the throughput of ARQ systems was treated which measures the transmission rate in $bit/symbol$ versus the time independent ratios $\frac{E_S}{N_0}$ or $\frac{E_b}{N_0}$. According to its definition, only transmitted packets, which are actually used for decoding, are utilized for the analysis. This proofed to be a useful definition since then, the troughput is independent of the transmission protocol and purely a function of the coding scheme integrated into the ARQ system and the channel. Hence, the integrated coding scheme could be compared on a fair basis to its theoretical limits, the channel capacity. For the end user, however, the throughput is of less concern. She or he is more interested in the temporal data rate of the system, measured in information bits per time. This performance measure is the topic of this chapter. In analogy with the throughput chapter, we define the data rate, derive its theoretical limits and finally analyze its dependency to previously defined performance measures and system parameters, schemes and protocols.

4.7.1 Definition

Due to the ARQ principle, the information bit rate is a random variable whose expected value will be called data rate R_{Data}:

> The data rate R_{Data} of digital communication system is defined as average number of transmitted information bits per time and measured in $\frac{bit}{sec}$.

Contrary to the definition of the throughput T on Page 72, there is no similar restriction to 'symbols required by the receiver'. Instead, the data rate is based on the time and consequently transmission gaps (as they occur in the SW protocol) and unused packets (as it is the case for the GBN protocol) lower the data rate. The on Page 93 defined average information delay $\overline{\Delta t}_I^{(N_I)}$ already reflected the influence of the protocol type for a performance measure. $\overline{\Delta t}_I^{(N_I)}$ represents the average time it takes to transmit N_I packets with L_{info} $bits$ each. This definition can be used to phrase the data rate definition mathematically: The average data rate is the limes

$$R_{Data} = \lim_{N_I \to \infty} \left\{ \frac{N_I \cdot L_{info}}{\overline{\Delta t}_I^{(N_I)}} \right\}. \tag{4.56}$$

In the next section, the channel capacity is used to derive theoretical limits for the data rate, whereas the section thereafter combines Equation 4.56 and the results of Section 4.5.3 to evaluate the data rates of ARQ systems.

4.7.2 Theoretical Limit

The channel capacity C relates the maximum throughput in $bit/symbol$ versus the signal energy to noise power density ratio $\frac{E_S}{N_0}$ or the energy per information bit to noise power density ration $\frac{E_b}{N_0}$. In order to derive the maximum possible data rate, we have to assume that all channel symbols are continuously sent, i.e. that there are no transmission gaps. Then, each symbol takes the time t_S and thus

$$R_{Data}^{max} = \frac{C}{t_S}. \tag{4.57}$$

For the one and two dimensional AWGN channel follows with Equations 4.22 and 4.23:

$$R_{Data}^{max,1D} = \frac{\log_2\left(1 + 2 \cdot \frac{E_S}{N_0}\right)}{2 \cdot t_S} \tag{4.58}$$

$$R_{Data}^{max,2D} = \frac{\log_2\left(1 + \frac{E_S}{N_0}\right)}{t_S}. \tag{4.59}$$

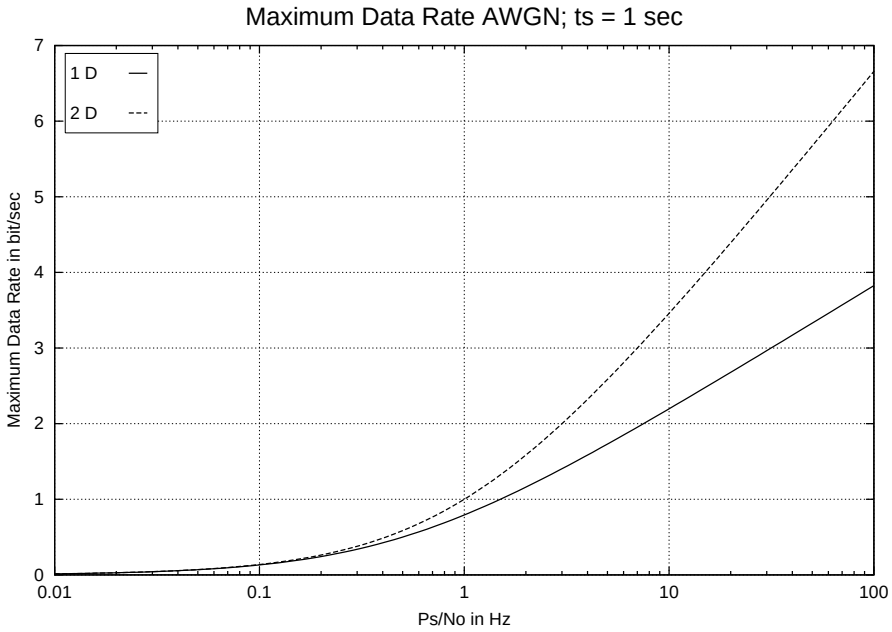

Figure 4.29: Maximum Data over 1-dim and 2-dim AWGN Channel with a Symbol Duration of $t_S = 1\ sec$.

The author likes to emphasis, that due to the division of the time-independent channel capacity with the symbol time, the resulting data rate is time dependent. In addition, the argument

$$\frac{E_S}{N_0} = \frac{P_S \cdot t_S}{N_0}$$

has also this symbol time incorporated. We therefore distinguish two special cases in the following

Constant Symbol Duration t_S

If we take the symbol duration t_S as a constant, the data rate becomes a function of the signal power to noise power density ratio $\frac{P_S}{N_0}$ for that particular symbol duration. This represents the case of a bandlimited ($t_S = const$) digital communication system, where variation in $\frac{P_S}{N_0}$ are due to a varying channel or a varying transmit power.

Figure 4.29 depicts the two maximum data rates for a symbol duration of $t_S = 1\ sec$. The plots are simply scaled versions of the corresponding channel capacities (compare Figure 4.29 with Figure 4.15 on Page 75). The larger the signal-power-to-noise-power-density ratio, the higher the effective data rate. Due to the fixed symbol duration, this plot are valid for that particular value only. Different symbol duration t_S, however, lead to different maximal rates for identical $\frac{P_S}{N_0}$ ratios. In general, the maximal data rate increases with smaller symbol times, since the numerators of Equations 4.58 and 4.59 are logarithmically dependent on t_S, whereas the denominators have a linear dependency. This behavior is exemplary depicted in the family of curves of Figure 4.30

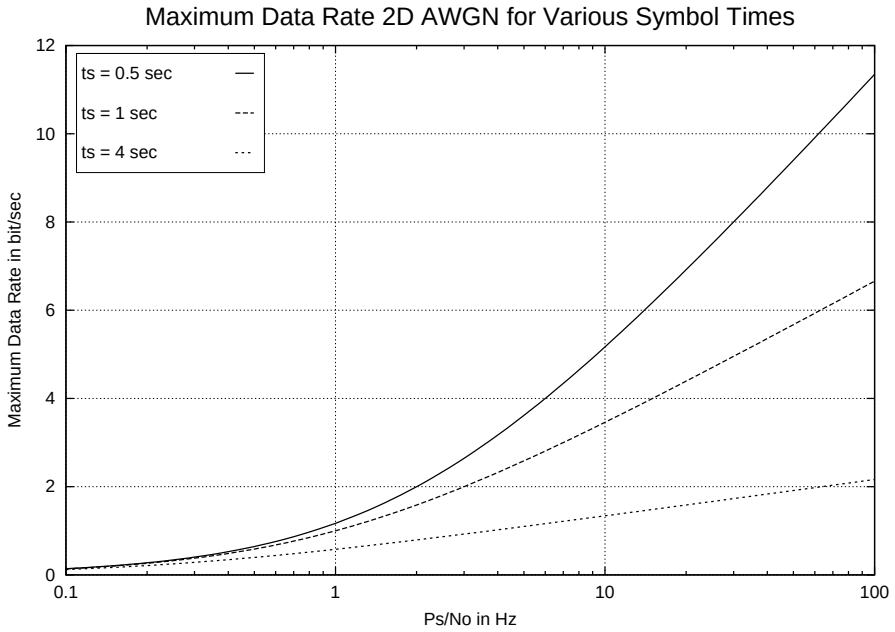

Figure 4.30: Maximum Data Rate over 2-Dim AWGN Channel for Different Symbol Durations.

for 3 different symbol durations. Although this behavior is obvious from its derivation, it seems strange that if the symbol duration and, hence the signal energy to noise power density ratio, is decreased the maximum possible data rate in fact increases. A reasoning can be found in the dimensionality of the sequences:

Take for example symbols sequence with symbol duration t_S transmitted over a 1D AWGN channel. If the symbol duration is split into half, the capacity is given by the 2D AWGN channel capacity at an effective 3 dB lower value. There are $\frac{E_S}{N_0}$ values where

$$C_{1D-AWGN}\left(\frac{E_S}{N_0}\right) < C_{2D-AWGN}\left(\frac{1}{2}\cdot\frac{E_S}{N_0}\right)$$

(see for example $\frac{E_S}{N_0} = 15\ dB$ in Figure 4.15 on Page 75), but there are also values for $\frac{E_S}{N_0}$ where

$$C_{1D-AWGN}\left(\frac{E_S}{N_0}\right) > C_{2D-AWGN}\left(\frac{1}{2}\cdot\frac{E_S}{N_0}\right)$$

(see for example $\frac{E_S}{N_0} = 0\ dB$ in Figure 4.15 on Page 75). Nevertheless, since the symbol times is cut into half, the data rate in $bit/time$ is twice that capacity at the lower value and the overall maximum data rate is with shortened symbol durations larger, i.e.

$$\frac{C_{1D-AWGN}\left(\frac{E_S}{N_0}\right)}{t_S} < \frac{C_{2D-AWGN}\left(\frac{1}{2}\cdot\frac{E_S}{N_0}\right)}{0.5\cdot t_S}.$$

Hence, as a rule it can be stated:

A digital communication system with constant signal-power-t- noise-power-density ratio and a reduced symbol duration of $\widetilde{t_S}$ has a higher theoretical maximal data rate $\widetilde{R}_{Data}^{max}$ than the corresponding system with symbol duration t_S and R_{Data}^{max}:

$$\widetilde{R}_{Data}^{max} > R_{Data}^{max} \; for \; \widetilde{t_S} < t_S.$$

However, to utilize this higher maximum data rate (at least partly) for a practical FEC communication system, one can not simply reduce the symbol duration. This indeed increases the data rate, but at the expanse of an increased error rate. Instead, some of the additional transmitted symbols must be used to lower the code rate until the same error rate is again achieved. Despite the lower code rate, the resulting data rate is still higher.

An ARQ system on the other hand, inherently adapts its data rate to the effectively changed $\frac{E_S}{N_0} = \frac{P_S}{N_0} \cdot t_S$. Hence, one expects that the shortening of the symbol duration automaticely leads to a higher data rate. In the Section 4.7.3, it will be shown that this is true for well designed ARQ systems, but not in general.

Constant Signal Power to Noise Power Density Ratio $\frac{P_S}{N_0}$

In Equations 4.58 and 4.59 the denominator as well as the numerator are dependent on the symbol duration t_S. Hence, if the signal power to noise power density ratio is held constant and the symbol duration is varied, the resulting maximal data rate curves are not simply scaled versions of the channel capacity as it was the previous discussed case for $t_S = const.$ Figure 4.31 depicts the maximum data rates for the 1D and the 2D AWGN channel with a constant signal power to noise power density ratio of $\frac{P_S}{N_0} = 1\,Hz$.

The curves reflect what was already discussed in the previous section: An increased symbol duration yields to a lower maximum possible data rate. As Figure 4.31 already suggests, even though the maximum data rate monotonically increases as the symbol duration decreases, the maximum data rate is not unlimited for a limited $\frac{P_S}{N_0}$. Instead the maximal value of R_{Data}^{max} for variations in t_S is achieved if the symbol duration approaches zero and equals, as shown in Appendix C.6.1, for both AWGN channels

$$\begin{aligned} \max_{t_S}\{R_{Data}^{max}\} &= \lim_{t_S \to 0^+}\{R_{Data}^{max}\} \\ &= \frac{1}{\ln(2)} \cdot \frac{P_S}{N_0}. \end{aligned}$$

In analogy to Figure 4.30, Figure 4.32 depicts a family of maximum data rate curves with the power to power density ratio $\frac{P_S}{N_0}$ as parameter. In this plot, also the dependency of the maximum data rate on the varying power to power density ratio can be seen: The larger this ratio, the higher the maximal data rate.

Using the Nyquist bandwidth

$$B_S = \frac{1}{t_S},$$

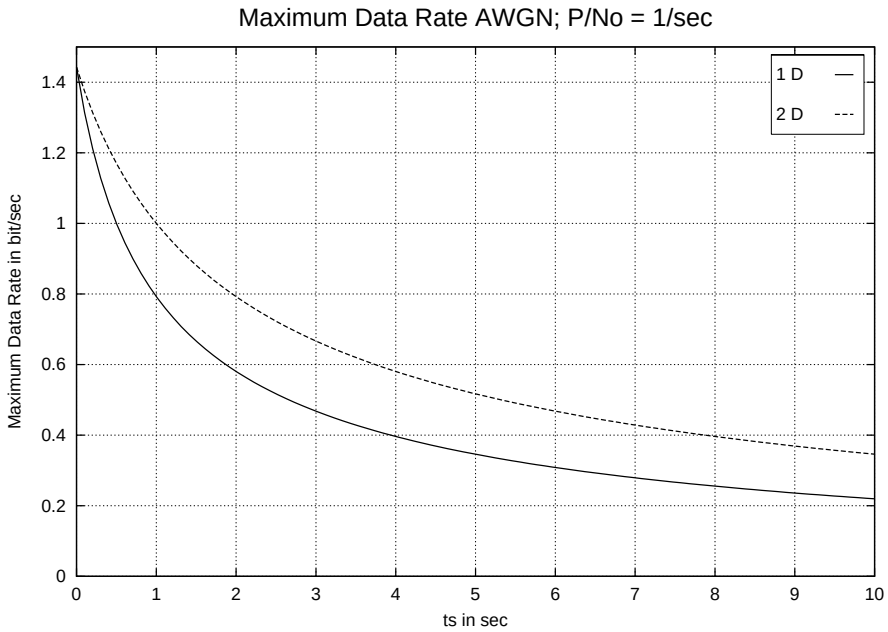

Figure 4.31: Theoretical Limit for Data Rate over the AWGN Channel with Signal-Power-to-Noise-Power-Density Ratio of $1/sec$.

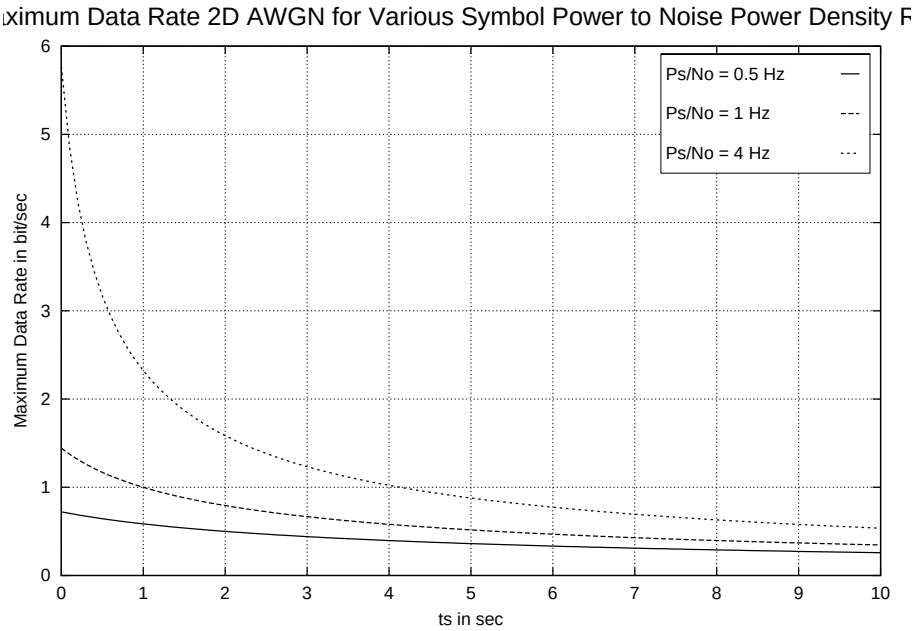

Figure 4.32: Maximum Data Rate over 2-Dim AWGN Channel for Different Signal Power to Noise Power Density Ratios.

the maximum data rate for constant $\frac{P_S}{N_0}$ can also be plotted versus that parameter.

4.7.3 Data Rate Performance

In previous section the maximum possible data rate over the AWGN channel was discussed. In analogy, this section begins with the derivation of the maximum possible data rate of an arbitrary ARQ system with throughput T and precedes to the derivation of the actual data rate performance of the ARQ system with the various transmission protocols. In order to do so, only the results of Section 4.5.3 and Equation 4.56 need to be combined.

Maximum Data Rate for a Given ARQ System

Equivalent with the derivation of the maximum possible data rate over an AWGN channel, the maximum data rate of a specific ARQ system with throughput T is obtained if all symbols are continuously sent and all sent symbols are used for a decoding attempt. Then, the maximum data rate $R_{Data}^{max,ARQ}$ in $bit/time$ is simply given be the division of the throughput T in $bit/symbol$ with the symbol duration t_S in $time/symbol$:

$$R_{Data}^{max,ARQ} = \frac{T}{t_S}. \tag{4.60}$$

Also in equivalence with the maximum data rate for the AWGN channel $R_{Data}^{max,AWGN}$, we distinguish the two special cases $t_S = const$ and $\frac{P_S}{N_0} = const$.

Constant Symbol Duration As Equation 4.60 already indicates, if the symbol duration t_S is constant, the maximum data rate of an ARQ system is simply a scaled version of the corresponding throughput T. This can be seen when Figure 4.33, which depicts the maximum data rate for the VE-MARQ system, is compared to the throughput of that system shown in Figure 4.21 on Page 82.

Constant Power to Noise Power Density Ratio As with the maximum AWGN channel data rate, if the symbol duration t_S is varied and the power to power density ratio $\frac{P_S}{N_0}$ is held constant, the data rate of an ARQ system is no scaled version of the throughput. Again, there will be a maximum data rate for $t_S \rightarrow 0^+$, but that limit can not be obtained analytically, since no analytical expression for T exists. As already mentioned, since for smaller symbol durations the maximum possible data rate increases and since an ARQ system automatically adapts its data rate to the channel conditions, we expect the actual data rate of an ARQ system to increase monotonically with smaller symbol durations. Figure 4.34, however reveals that this is not true for the considered system. There are some symbol duration regions (e.g. at $t_S = 3\ sec$ and $t_S = 0.4sec$) where a lower symbol duration also yields a lower maximum data rate. The existence of such regions again indicates bad system design. To be more specific, these symbol duration

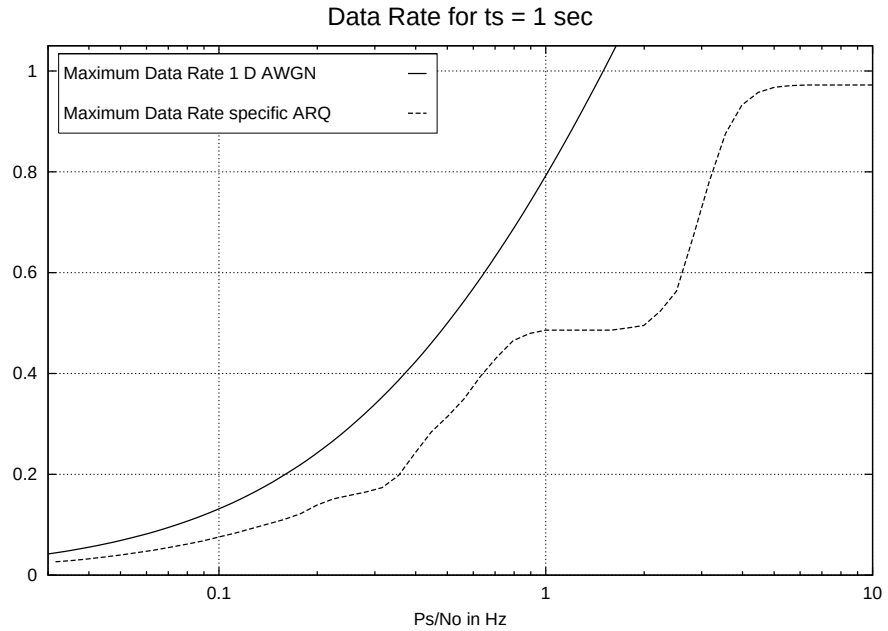

Figure 4.33: Maximum Data Rate of the VE-MARQ System with Average Number of Transmissions of Figure 4.13 (Page 71) and Throughput of Figure 4.21 (Page 82) for Constant Symbol Duration t_S.

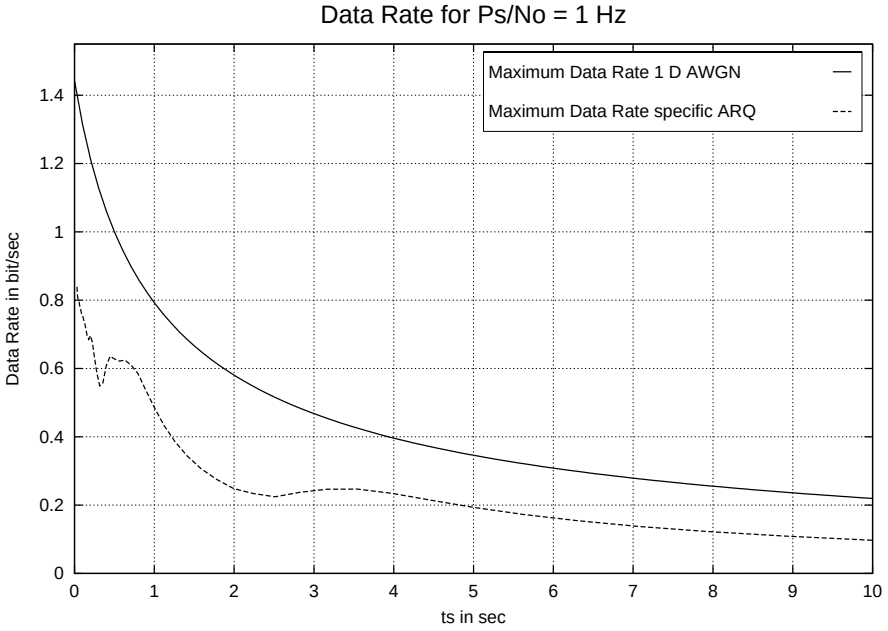

Figure 4.34: Maximum Data Rate of the VE-MARQ System with Average Number of Transmissions of Figure 4.13 (Page 71) and Throughput of Figure 4.21 (Page 82) for Constant Power to Power Density Ratio $\frac{P_S}{N_0}$.

regions directly correspond to regions in the throughput versus $\frac{E_b}{N_0}$ plot (see Figure 4.18 on Page 78) where an increase in $\frac{E_b}{N_0}$ yields a lower throughput.

As an example for such regions consider the throughput region $T \in [0.55, 0.82]$ in Figure 4.18. From the throughput versus $\frac{E_S}{N_0}$ plot in Figure 4.21 on Page 82 we obtain the corresponding signal energy to noise power density region $\frac{E_S}{N_0} \in [3.8\,dB, 5.8\,dB]$ which relates for $\frac{P_S}{N_0} = 1\,Hz$ into the symbol duration region $t_S \in [2.4\,sec, 3.8\,sec]$.

Hence, the maximum data rate versus symbol duration plot reveals, as the throughput vs. $\frac{E_b}{N_0}$ plot, bad system design and has the advantage of being much easier obtainable than the latter one.

Stop-and-Wait Protocol

As mentioned, to derive the data rate performance of an ARQ system, only the results of Section 4.5.3 and Equation 4.56 need to be combined. Hence, substituting Equation 4.48 into Equation 4.56 yields

$$\begin{aligned} R_{Data}^{SW} &= \lim_{N_I \to \infty} \left\{ \frac{N_I \cdot L_{info}}{t_{RT} \cdot N_I \cdot \left(\overline{n_{trans}} + \frac{1}{N_{RT}} \cdot \frac{1}{T} \right)} \right\} \\ &= \lim_{N_I \to \infty} \left\{ \frac{L_{info}}{t_{RT} \cdot \overline{n_{trans}} + t_{RT} \cdot \frac{1}{N_{RT}} \cdot \frac{1}{T}} \right\}, \end{aligned}$$

and the limit variable N_I cancels. Accordingly,

$$\begin{aligned} R_{Data}^{SW} &= \frac{L_{info}}{t_{RT} \cdot \overline{n_{trans}} + L_{info} \cdot t_S \cdot \frac{1}{T}} \\ &= \frac{1}{t_S} \cdot \frac{1}{\frac{t_{RT}}{t_S \cdot L_{info}} \cdot \overline{n_{trans}} + \frac{1}{T}} \end{aligned} \tag{4.61}$$

and finally,

$$\begin{aligned} R_{Data}^{SW} &= \frac{T}{t_S} \cdot \frac{1}{N_{RT} \cdot \overline{n_{trans}} \cdot T + 1} \\ &= R_{Data}^{max,ARQ} \cdot \frac{1}{N_{RT} \cdot \overline{n_{trans}} \cdot T + 1} \\ &= R_{Data}^{max,ARQ} \cdot K^{SW} \end{aligned} \tag{4.62}$$

Equation 4.62 shows that the average data rate of a general ARQ system with SW protocol depends on the maximum possible data rate with that particular ARQ system and a second term

$$K^{SW} = \frac{1}{N_{RT} \cdot \overline{n_{trans}} \cdot T + 1}, \tag{4.63}$$

which is always smaller than one, representing the degradation of the data rate due to the SW protocol. The protocol degradation term K^{SW} is for systems with uneven packet sizes not constant with varying $\frac{E_S}{N_0}$ since the product $\overline{n_{trans}} \cdot T$ is not constant: For high signal to noise ratios,

$\overline{n_{trans}} = 1$ and $T = R_{CM}^{FEC,1}$ and hence

$$\begin{aligned}\lim_{\frac{E_S}{N_0}\to\infty}\left\{K^{SW}\right\} &= \frac{1}{N_{RT}\cdot 1\cdot R_{CM}^{FEC,1}+1}\\ &= \frac{1}{n_{RT}^{(1)}+1}.\end{aligned}$$

The limit of K^{SW} for $\frac{E_S}{N_0} \to -\infty\, dB$, or equivalently $n_{trans} \to \infty$, is

$$\lim_{\frac{E_S}{N_0}\to-\infty}\left\{K^{SW}\right\} = \lim_{n_{trans}\to\infty}\left\{\frac{1}{N_{RT}\cdot n_{trans}\cdot R_{CM}^{ARQ,n_{trans}}+1}\right\}$$

The value of this limit depends on the sequence of coding-/modulation rates. However, any practical ARQ system has from a certain number of transmissions $\widetilde{n}_{trans}$ on a constant packet size $L_{\widetilde{n}_{trans}}$ and hence a constant FEC coding-/modulation rate $R_{CM}^{FEC,\widetilde{n}_{trans}}$ for these packets. In Appendix C.6.2 it is shown that in this case

$$\begin{aligned}\lim_{\frac{E_S}{N_0}\to-\infty}\left\{K^{SW}\right\} &= \frac{1}{N_{RT}\cdot R_{CM}^{FEC,\widetilde{n}_{trans}}+1}\\ &= \frac{1}{n_{RT}^{\left(\widetilde{n}_{trans}\right)}+1}.\end{aligned}$$

In general, the protocol degradation factor K^{SW} has a maximum variation of

$$\frac{1}{n_{RT}^{max}+1} \le K^{SW} \le \frac{1}{n_{RT}^{min}+1}$$

and as a consequence it is almost constant if N_{RT} is small (then the variation of n_{RT} has a small influence in the variation of K^{SW}) or if the packets sizes hardly vary.

Figure 4.35 shows the product $\overline{n_{trans}}\cdot T$ and K^{SW} for the VE-MARQ system with coding-/modulation rate sequence of $R_{CM}^{ARQ,j} = \frac{280}{288}\cdot\left\{1,\frac{1}{2},\frac{1}{3},\frac{1}{6},\ldots,\frac{1}{3\cdot(j-2)},\ldots\right\}$ and the two sample environments defined in Section 4.5.1. The product $\overline{n_{trans}}\cdot T$ varies between $\frac{280}{288}\cdot\frac{1}{3}$ and $\frac{280}{288}$ and as a consequence the SW degradation factor from 0.0382 to 0.1065 for the satellite system and from 0.888 to 0.960 for the terrestrial sample system.

If the **packet size is constant** the SW protocol data rate degradation factor can be simplified to

$$\begin{aligned}K^{SW} &= \frac{1}{N_{RT}\cdot R_{CM}^{FEC}+1}\\ &= \frac{1}{n_{RT}+1}.\end{aligned}$$

Hence, the data rate of an ARQ system with constant packet sizes and SW protocol is a constant fraction of the maximum possible data rate $R_{Data}^{max,ARQ}$ with that particular ARQ system.

Independent of the fact, whether we have a system with constant packets or not, large val-

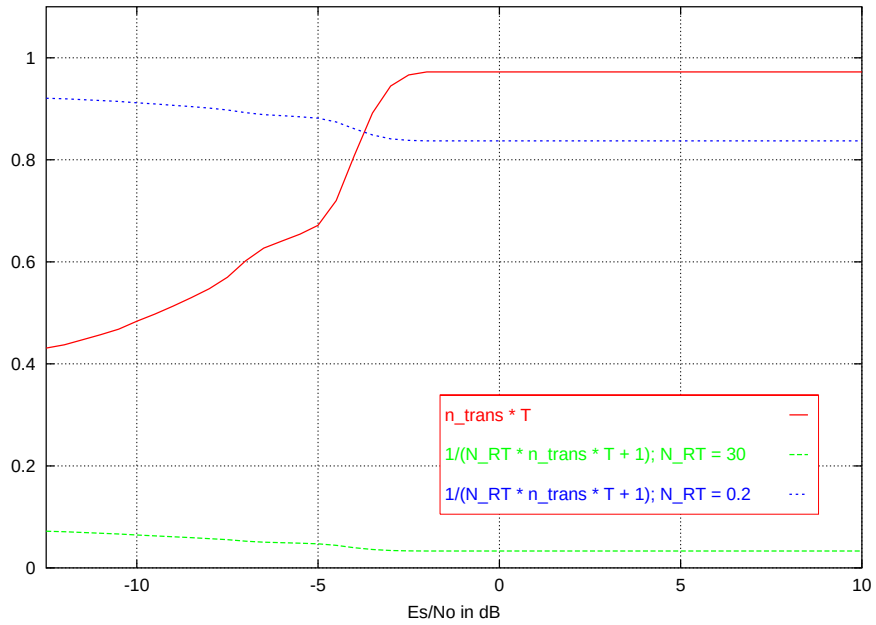

Figure 4.35: Protocol Degradation Factor K^{SW} for the Two Sample Environments and the Product $\overline{n_{trans}} \cdot T$.

ues for the round trip number N_{RT} in Equation 4.63 yields to very small values for K^{SW} and consequently very small data rates. As a general rule it can be stated

> The Stop-and-Wait protocol does not lead to extreme data rate performance degradation only if the round trip delay is a small fraction of the packet durations.

Therefore, the SW protocol should be avoided in systems with a round trip number exceeding 0.1. Figure 4.36 shows the degradation factor K^{SW} as a function of the number of packets on the round trip route n_{RT} for systems with constant packet sizes.

As with the maximum data rate, if we want to plot the data rate of an ARQ system, we have to decide which parameter shall be held constant, t_S or $\frac{P_S}{N_0}$.

Constant Symbol Duration t_S If the symbol duration is held constant, the maximum possible data rate with SW protocol is for high signal to noise ratios:

$$\begin{aligned} R_{Data}^{max.SW} &= \frac{T}{t_S} \cdot \frac{1}{n_{RT}^{(1)}+1} \\ &= \frac{T}{t_S} \cdot \frac{t_1}{t_{RT}+t_1}. \end{aligned}$$

The data rate at an arbitrary $\frac{E_S}{N_0}$ value can be obtained from Equation 4.62. Figure 4.37 shows data rates R_{Data} versus the signal power to noise power density ratio ($t_S = const$) plots for the VE-MARQ system with the average number of transmissions depicted in Figure 4.13 on Page

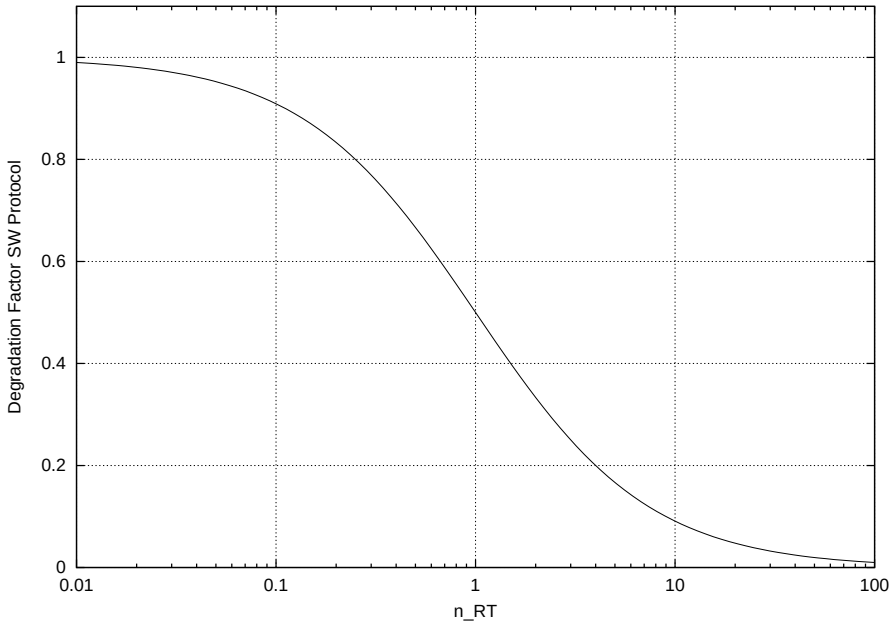

Figure 4.36: Degradation Factor of Stop-and-Wait Protocol versus the Number of Packets on the Round Trip Route.

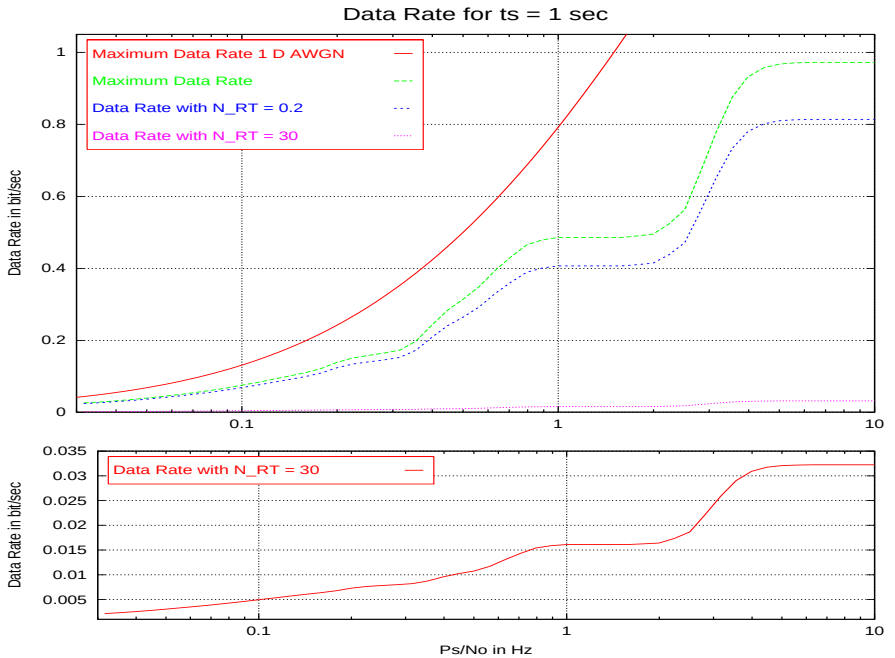

Figure 4.37: Data Rate of the VE-MARQ System with Average Number of Transmissions of Figure 4.13 (Page 71) and Throughput of Figure 4.21 (Page 82) with SW Protocol for Constant Symbol Duration t_S.

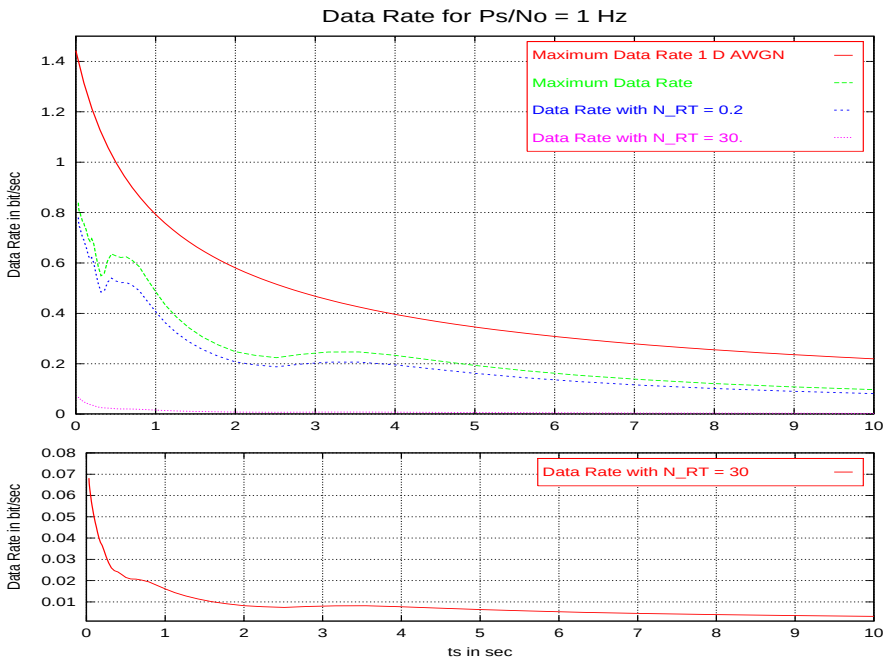

Figure 4.38: Data Rate of the HARQ System with Average Number of Transmissions of Figure 4.13 (Page 71) and Throughput of Figure 4.21 (Page 82) with SW Protocol for Constant Signal Power to Noise Power Density Ratio $\frac{P_S}{N_0}$.

71 and throughput depicted in Figure 4.21 on Page 82 with SW protocol in comparison with the maximum possible 1D AWGN data rate and the maximum possible data rate for that particular ARQ systems.

What stands out is the low data rate of the system using the satellite link - as mentioned a consequence of the large round trip number. Therefore, the lower plot provides an enlargement of this data rate.

Constant Signal Power to Noise Power Density Ratio The protocol degradation factor K^{SW} is, as we discussed, only dependent on the overall change of $\frac{E_S}{N_0}$, independent of whether t_S or $\frac{P_S}{N_0}$ was actually responsible for that change. Consequently, the data rate versus t_S curve is like the data rate versus $\frac{P_S}{N_0}$ curve a by the protocol degradation factor K^{SW} scaled version of the corresponding maximum data rate curve for that ARQ system.

Figure 4.38 depicts the corresponding data rate plot to Figure 4.37, but this time versus t_S.

Go-Back-N Protocol

As for the SW protocol, substitution of Equation 4.49 into Equation 4.56 yields for the data rate of an ARQ system with GBN protocol

$$\begin{aligned} R_{Data}^{GBN} &= \lim_{N_I\to\infty}\left\{\frac{N_I\cdot L_{info}}{t_{RT}\cdot\left(N_I\cdot\overline{n_{trans}}-N_I-1+\frac{N_I}{N_{RT}}\cdot\frac{1}{T}\right)}\right\} \\ &= \lim_{N_I\to\infty}\left\{\frac{L_{info}}{t_{RT}\cdot\left(\overline{n_{trans}}-1-\frac{1}{N_I}+\frac{1}{N_{RT}}\cdot\frac{1}{T}\right)}\right\} \\ &= \frac{L_{info}}{t_{RT}\cdot\left(\overline{n_{trans}}-1+\frac{1}{N_{RT}}\cdot\frac{1}{T}\right)}. \end{aligned}$$

With Equation 4.37 we obtain

$$\begin{aligned} R_{Data}^{GBN} &= \frac{L_{info}}{t_{RT}(\overline{n_{trans}}-1)+\frac{t_S\cdot L_{info}}{T}} \\ &= \frac{1}{t_S}\cdot\frac{1}{\frac{t_{RT}}{t_S\cdot L_{info}}\cdot(\overline{n_{trans}}-1)+\frac{1}{T}} \end{aligned}$$

and finally,

$$\begin{aligned} R_{Data}^{GBN} &= \frac{T}{t_S}\cdot\frac{1}{N_{RT}\cdot(\overline{n_{trans}}-1)\cdot T+1} \\ &= R_{Data}^{max,ARQ}\cdot K^{GBN}. \end{aligned} \tag{4.64}$$

Again, the data rate can be written as the product of the maximum data rate with a protocol degradation factor. This factor is for the GBN protocol

$$K^{GBN} = \frac{1}{N_{RT}\cdot(\overline{n_{trans}}-1)\cdot T+1}. \tag{4.65}$$

The discussion of the data rate of a system with GBN protocol can be limited to the discussion of this degradation factor, since the maximum data rate for an ARQ system was already treated in the beginning of this section.

Again, the lowest degradation is obtained if no retransmissions occur, i.e. $\overline{n_{trans}} = 1$ and $T = R_{CM}^{FEC,1}$. For the GBN protocols it is

$$\begin{aligned} \lim_{\frac{E_S}{N_0}\to\infty}\left\{K^{GBN}\right\} &= \frac{1}{N_{RT}\cdot(1-1)\cdot R_{CM}^{FEC,1}+1} \\ &= 1 \end{aligned}$$

and hence no degradation from the maximum possible data rate occurs for excellent channel conditions. For $\frac{E_S}{N_0}\to-\infty$ the limit value of K^{GBN} can also not be obtained without further assumptions. However, if the system uses constant packet sizes with a FEC coding-/modulation rate $R_{CM}^{FEC,\widetilde{n}_{trans}}$ with the beginning of the $\widetilde{n}_{trans}$-th transmission we obtain , since the term $N_{RT}\cdot T$ cancels, the same limit value as for the SW protocol

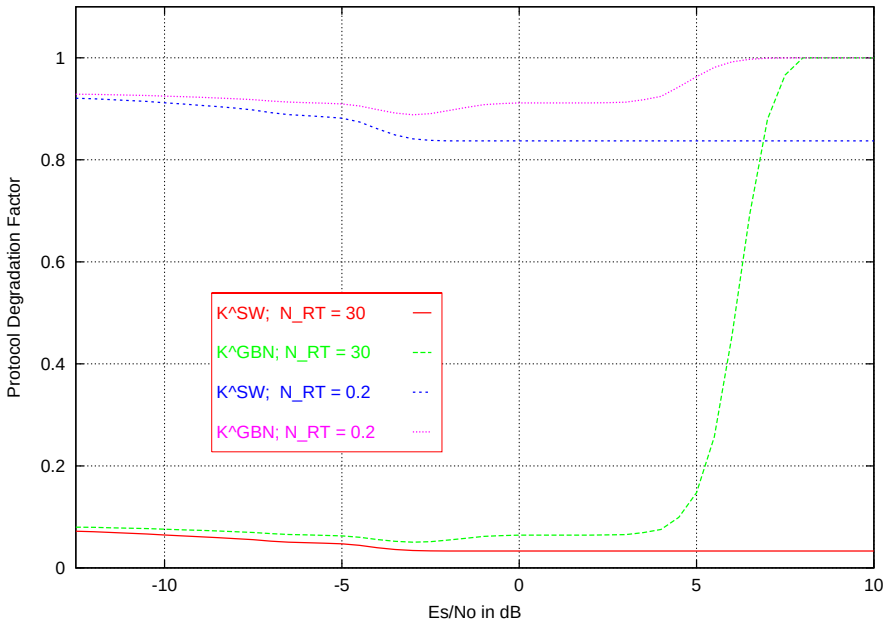

Figure 4.39: Protocol Degradation Factors K^{SW} and K^{GBN} for the Two Sample Environments.

$$\lim_{\frac{E_S}{N_0}\to-\infty}\left\{K^{GBN}\right\} = \frac{1}{N_{RT}\cdot R_{CM}^{FEC,\widetilde{n_{trans}}}+1}$$
$$= \frac{1}{n_{RT}^{\left(\widetilde{n_{trans}}\right)}+1}.$$

This two limits already reveal that the degradation factor K^{GBN} has a much greater dynamic range than K^{SW}, especially for systems with a large round trip number. Independent of the environment type, there is no degradation for good channels. As soon as the channel worsens, however, the degradation factor asymptotically approaches the degradation of the SW protocol. Figure 4.39 illustrates this effect. As general rules it can be stated

> The Go-Back-N protocol provides a noticeable data rate improvement compared to the Stop-and-Wait protocol only of there are hardly any retransmissions.
>
> The Go-Back-N protocol also yields to unacceptable data rate degradations for systems with large round trip numbers as soon as retransmissions occur.

Contrary to the SW protocol, also system with constant packet sizes have a varying degradation factor. More specifically, for these systems holds

$$K^{GBN} = \frac{1}{n_{RT} - N_{RT}\cdot T + 1}$$

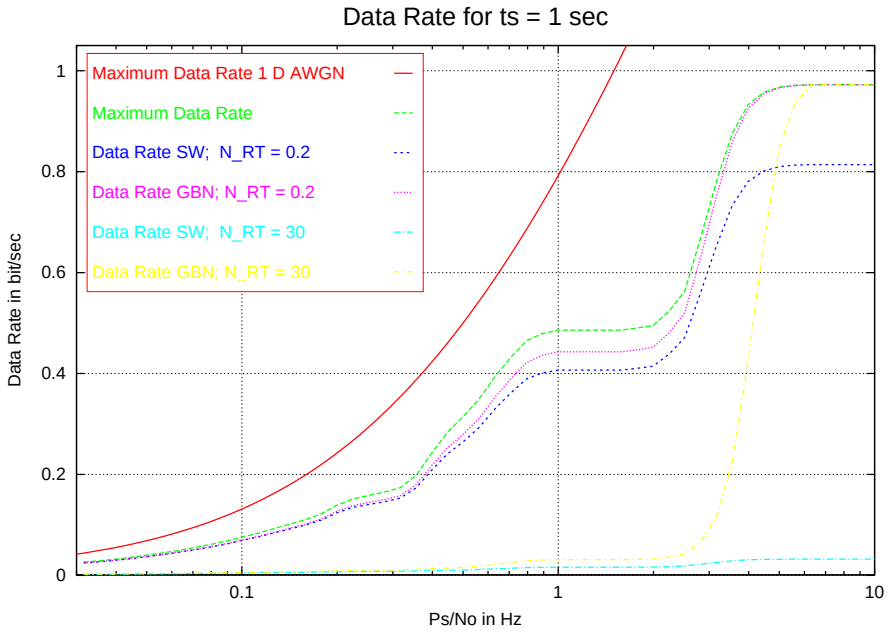

Figure 4.40: Data Rates the VE-MARQ System with Average Number of Transmissions of Figure 4.13 (Page 71) and Throughput of Figure 4.21 (Page 82) with SW and GBN Protocol for Constant Symbol Duration t_S.

and the degradation factor monotonically varies between

$$\frac{1}{n_{RT}+1} \leq \frac{1}{n_{RT} - N_{RT} \cdot T + 1} \leq 1.$$

Hence, a corresponding plot to Figure 4.36, which is independent of the throughput of the ARQ system can not be given.

Figure 4.40 depicts the resulting data rates of the example ARQ system with GBN and SW protocol if again the symbol duration t_S is held constant. Everything, which was said about the degradation factor, is reflected in the data rate. For the terrestrial system, the GBN protocol yields a small data rate gain in the region where there are only a few retransmissions. For the satellite system, however, the GBN protocol provides a huge gain when there are no retransmissions at all, marginal gain as soon as retransmissions occur.

Figure 4.41 shows the corresponding data rate plot if the signal power to noise power density is held constant. Again, the GBN protocol provides a small gain for the terrestrial system and for the satellite system only a gain where no retransmissions occur.

Selective Repeat Protocol

In the derivation of the maximum data rate for a given ARQ system with throughput T it was stated, that it is achieved only if all symbols are sent continuously and all symbols are used for a decoding attempt. This is, however, exact the way the ideal SR protocol works. Hence it can be

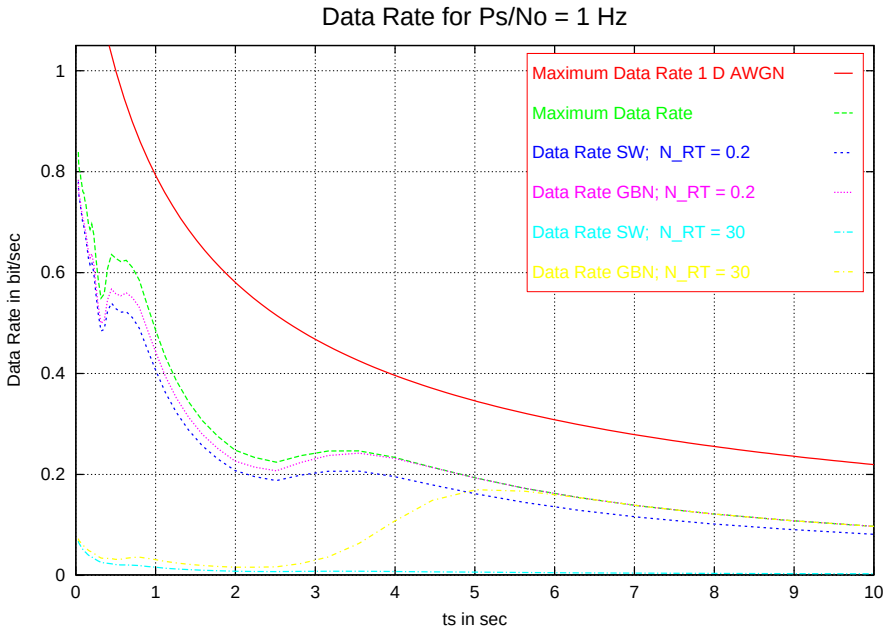

Figure 4.41: Data Rates of the VE-MARQ System with Average Number of Transmissions of Figure 4.13 (Page 71) and Throughput of Figure 4.21 (Page 82) with SW and GBN Protocol for Constant Signal Power to Noise Power Density Ratio $\frac{P_S}{N_0}$.

concluded that the SR protocol has the maximum possible data rate independent of the number of transmissions:

$$R_{Data}^{SR} = R_{Data}^{max,ARQ}.$$

This result can also be obtained mathematically if the bound of the average information delay (Bounds 4.51) are substituted in Equation 4.56:

$$\begin{aligned} R_{Data}^{SR} &= \lim_{N_I \to \infty} \left\{ \frac{N_I \cdot L_{info}}{t_{RT} \cdot \left(C_1 + \frac{N_I}{N_{RT}} \cdot \frac{1}{T} \right)} \right\} \\ &= \lim_{N_I \to \infty} \left\{ \frac{L_{info}}{t_{RT} \cdot \left(\frac{C_1}{N_I} + \frac{1}{N_{RT}} \cdot \frac{1}{T} \right)} \right\} \\ &= \frac{L_{info}}{\frac{t_S \cdot Linfo}{T}} \end{aligned}$$

and hence

$$R_{Data}^{SR} = \frac{T}{t_S}.$$

The corresponding performance degradation factor is consequently

$$K^{SR} = 1.$$

Since the data rate of a system with SW protocol does not deviate from the maximum possible data rate, Figures 4.40 and 4.41 can be used for a comparison of this protocol type with the two others.

Chapter 5

Diversity Combining Techniques

Throughout the preceeding chapter, ARQ performance measures were defined and their mutual relation were derived. The sequence of retransmission probabilities $P(RR_j)$ and their bounds, composed of the rejection probabilities $P(R_j)$, served as a starting point for these series of derivations. All other ARQ performance measures were derived with the help of these probabilities and other system and environment parameters. This whole treatment was totally independent of the type of ARQ system (CE or VE-ARQ) and its realization (with or without memory). Of course, the actual realization of an ARQ system and the used channel model inherently determines exactly this set of retransmission probabilities.

In Chapter 4 already some actual ARQ realizations served as illustration. Memoryless systems were found to be highly inefficient (see for example Figure 4.11). If memory is incorporated into the ARQ system, the question arises what kind of redundant information shall be sent for retransmissions. The straight forward approach relies on repetitions, leading to a CE-MARQ system. A more efficient approach would send retransmissions, which can be used to successively construct better codes. A system with this behavior could, for example, be constructed via puncturing of a convolutional code. In this case, retransmissions are code bits which were punctured and not transmitted previously. This strategy can be continued until the mother code is completely transmitted. If this is achieved and the decoding process is still not successful, even more additional redundancy is required.

This problem will be the topic of the present chapter. If further redundancy is required and one decides to send repetitions of the mother codeword, how are all these transmissions properly combined for the introduced channel models? Also, if one derives from the same mother codeword a different channel word via changing the mapper, how are these retransmissions properly combined? And does this mapper change lead to an improvement?

In order to answer these question, the chapter is divided into two parts. The first one treats the mentioned questions for codes which use maximum likelihood decoding, such as convolutional codes. For a combination of these codes with multilevel modulation (TCM), Schmitt already presented an analysis for the AWGN channel in [Schm98]. In this section, these results are ex-

tended to all channel models of Chapter 3. Also in [Schm98][Schm99] a simple way of deriving successive channel sequences from the same mother codeword for multilevel modulation was described. The first part of this chapter is concluded by a review of these results. The second part of the chapter is devoted to combining for codes which make use of so-called maximum a posteriori decoding (MAP), such as turbo codes. Proper combining for these codes is investigated and in analogy with the ML section a novel technique to derive a new channelword from the mother codeword for retransmission is presented. Results for this technique conclude this chapter.

5.1 Maximum Likelihood Combining

As mentioned above, this section is concerned with the problem how to properly combine several retransmissions of a codeword with possibly varying mapper, if the code is to be decoded by a maximum likelihood (ML) decoder. In general, ML decoding is concerned with the selection of the codesequence from the set of all possible code sequences, which makes the channel outcome most likely. Therefore, the so-called likelihood function needs to be maximized. Section 5.1.1 derives the likelihood function for all presented channel models if only a single tranmission is available. Hence, this corresponds to the case of FEC error control. The following section also derived the likelihood function for all introduced channel models, however, we assumes that several transmissions (the first and a certain number of retransmissions) are available. During these derivations we allow for varying mapper functions for the individual transmissions.

Based on these results, Section 5.1.3 determines how multiple transmissions of a packet can be combined for maximum likelihood decoding.

It will be seen that for ARQ schemes with a constant encoder and mapper a symbolwise combining exists, which also enables maximum likelihood decoding. This so-called maximum-ratio-combining (MRC) will be investigated in Section 5.1.4. Finally, this section is concluded with review of the performance gains with MRC and additional gains, which can be obtained if the encoder mapping is varied for the retransmissions.

5.1.1 Maximum Likelihood Decoding of Single Transmissions

Figure 3.1 on Page 22 depicted the generic transmission process of a digital communication system. From the many digital operations, which come to application in order to prepare the information for transmission, only the encoding and the mapping process are of concern for this work. Equivalently, at the receiver side only the reverse processes metric generation from the already soft demodulated[1] receive symbols and their decoding are of interest. If these blocks are substituted for the digital operation blocks in Figure 3.1 and if the continuous channel is replaced with its discrete model the block diagram of Figure 5.1 is obtained. Herein $\underline{b} = (b[0], b[1], \ldots, b[k], \ldots, b[L_{info}-1])$ represents the sequence of bits to be transmitted and $\underline{c} = (c[0], c[1], \ldots, c[k], \ldots, c[L-1])$ the

[1] Contrary to hard demodulation, the demodulator does not decide for a certain channel symbol, but resolves the analog signal into its basis functions. The output of a soft demodulator are the corresponding I and Q values.

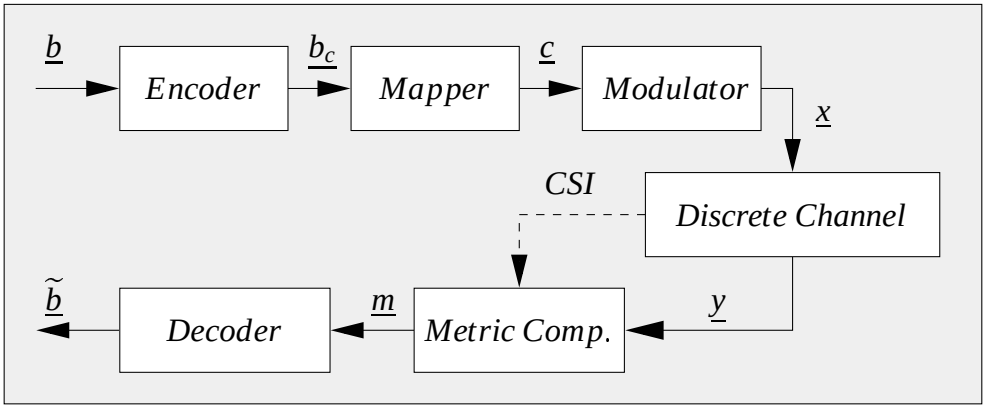

Figure 5.1: Simplified Block Diagram of Digital Transmission with Discrete Channel Model

so-called code symbol sequence. The code symbol sequence is obtained via grouping a certain amount of encoder output bits in order to match the modulation alphabet. Hence, the lengths L_{info} and L of the information bit and the code symbol sequences are related by the coding-/modulation rate $R_{CM} = \frac{L_{info}}{L}$, as discussed in Section 4.1. The encoder feeds his output symbols into the mapper which maps them memoryless onto channel symbols $x[k]$ from the alphabet $X = \{x_0, x_1, \ldots, x_{M-1}\}$, $x_j \in \mathbf{C}$ to form the complex channel symbol sequence $\underline{x} = (x[0], x[1], \ldots, x[k], \ldots x[L-1]) \in X^L$.

As an example, consider the case of a rate $\frac{1}{4}$ code with an information word length of $100\ bit$ which is transmitted via QPSK modulation. Then, the encoder output consists of $400\ bit$, which have to be grouped into bit pairs ($M = 4$) resulting in a code symbol sequence $\underline{c}$ and a channel symbol sequence $\underline{x}$ of length $200\ symbol$ and a coding-/modulation rate of $R_{CM} = \frac{1}{2}$.

The transmitted channel symbol sequence is corrupted by the discrete channel, resulting in the receive sequence $\underline{\mathbf{y}} = (\mathbf{y}[0], \mathbf{y}[1], \ldots, \mathbf{y}[k], \ldots \mathbf{y}[L-1]) \in \mathbf{C}^L$. This sequence as well as the channel state information (CSI; see Figure 5.1) is used to generate a sequence of metrics $\underline{m}$ on which the decoder bases its decoding process, finally leading to an estimate

$$\underline{\widetilde{b}} = \left(\widetilde{b}[0], \widetilde{b}[1], \ldots, \widetilde{b}[k], \ldots, \widetilde{b}[L_{info}-1]\right)$$

for the transmitted bit sequence $\underline{b}$.

In the decoding process, the maximum likelihood (ML) decoder decides for one of the $2^{L_{info}}$ possible channel symbol sequences $\underline{x}$ as an estimate $\underline{\widetilde{x}}$ for the transmitted channel symbol sequence. This selection is based on the principle, that a decision is made for the one, which makes the observation of the receive sequence most likely (hence the name). That means in case of hard decision decoding, that the decision is made for the code sequence $\underline{\widetilde{c}}$ which maximizes the conditioned probability $P_{\underline{\mathbf{y}^H}|\underline{c}}\left(\underline{y^H}\middle|\,\underline{\widetilde{c}}\right)$, where $\underline{y^H}$ respresents the hard decision demodulated receive sequence. Equivalently, in case of soft decision, the decoder decides for the channel symbol se-

quence $\underline{x}$, which maximizes the conditioned probability density function

$$f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y}\middle|\,\underline{x}\right). \tag{5.1}$$

This conditioned PDF is the so-called likelihood function for the receive sequence $\underline{\mathbf{y}} = \underline{y}$. Due to the monotonically increasing behavior of the logarithm function, maximizing Equation 5.2 is identical to maximizing its logarithm, the so-called log-likelihood function

$$ln\left(f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y}\middle|\,\underline{x}\right)\right).$$

As a result of the one-to-one mapping of the encoder, the decision for a channel symbol sequence automatically implies a decision for the corresponding information sequence $\underline{\widetilde{b}}$ as estimate for the transmitted information sequence $\underline{b}$.

In the remainder of this section, the likelihood and the log-likelihood functions for a single receive sequence $\underline{y}$ and the in Chapter 3 presented channel models are derived.

AWGN Channel

All channel models, presented in Chapter 3, had in common that the additive noise parts were symbolwise mutually independent. Consequently, for the AWGN channel, where the additive noise is the only corruption of the transmitted sequence, the likelihood function 5.1 can be written as multiplication of the conditioned PDFs $f_{\mathbf{y}|x}\left(y\left[k\right]\middle|\,x\left[k\right]\right)$ of the individual symbols

$$f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y}\middle|\,\underline{x}\right) = \prod_{k=0}^{L-1} f_{\mathbf{y}|x}\left(y\left[k\right]\middle|\,x\left[k\right]\right) \tag{5.2}$$

and the log-likelihood function is given by the sum

$$ln\left(f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y}\middle|\,\underline{x}\right)\right) = \sum_{k=0}^{L-1} ln\left(f_{\mathbf{y}|x}\left(y\left[k\right]\middle|\,x\left[k\right]\right)\right). \tag{5.3}$$

In order to further resolve the likelihood and the log-likelihood functions, more details about the underlying channel model are required.

If the **AWGN channel model 1** (Figure 3.2 on Page 26) is used, substitution of Equation 3.6 yields

$$f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y}\middle|\,\underline{x}\right) = \frac{1}{\left[\pi \cdot N_0\right]^L} \cdot e^{-\frac{1}{N_0}\cdot\sum_{k=0}^{L-1}|y[k]-x[k]|^2} \tag{5.4}$$

and for the log-likelihood function follows

$$\begin{aligned} ln\left(f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y}\middle|\,\underline{x}\right)\right) &= -\tfrac{1}{N_0}\cdot\textstyle\sum_{k=0}^{L-1}\left|y\left[k\right]-x\left[k\right]\right|^2 - L\cdot ln\left(\pi\cdot N_0\right) \\ &= -K_1\cdot\textstyle\sum_{k=0}^{L-1}\left|y\left[k\right]-x\left[k\right]\right|^2 - K_2 \end{aligned}. \tag{5.5}$$

Since $K_1 > 0$, finding the channel symbol sequence $\underline{x}$, which maximizes the log-likelihood function 5.5, is identical with finding the sequence $\underline{x}$, which minimizes the squared Euclidean distance $d_E^2\left(\underline{x}, \underline{y}\right)$ to the receive sequence $\underline{\mathbf{y}} = \underline{y}$:

$$\begin{aligned} d_E^2\left(\underline{x}, \underline{y}\right) &= \sum_{k=0}^{L-1} \left|y\left[k\right] - x\left[k\right]\right|^2 \\ &= \sum_{k=0}^{L-1} d_E^2\left(x\left[k\right], y\left[k\right]\right) \end{aligned} \tag{5.6}$$

If, on the other hand, **AWGN model 2** (Figure 3.3 on Page 27) is used, the likelihood function

$$f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y}\middle|\,\overline{\underline{x}}\right) = \left[\frac{E_S}{\pi \cdot N_0}\right]^L \cdot e^{-\frac{E_S}{N_0} \cdot \sum_{k=0}^{L-1} |y[k] - \overline{x}[k]|^2} \tag{5.7}$$

and the log-likelihood function

$$\begin{aligned} ln\left(f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y}\middle|\,\overline{\underline{x}}\right)\right) &= -\tfrac{E_S}{N_0} \cdot \sum_{k=0}^{L-1} \left|y\left[k\right] - \overline{x}\left[k\right]\right|^2 + L \cdot ln\left(\tfrac{E_S}{\pi \cdot N_0}\right) \\ &= -K_3 \cdot \sum_{k=0}^{L-1} \left|y\left[k\right] - \overline{x}\left[k\right]\right|^2 + K_4 \end{aligned} \tag{5.8}$$

are obtained. Applying the same argumentation as for AWGN model 1 yields that ML decoding is identical to choosing the valid channel symbol sequence $\overline{\underline{x}}$, which also minimizes the squared Euclidean distance

$$\begin{aligned} d_E^2\left(\overline{\underline{x}}, \underline{y}\right) &= \sum_{k=0}^{L-1} \left|y\left[k\right] - \overline{x}\left[k\right]\right|^2 \\ &= \sum_{k=0}^{L-1} d_E^2\left(\overline{x}\left[k\right], y\left[k\right]\right) \end{aligned} \tag{5.9}$$

to the receive sequence $\underline{\mathbf{y}} = \underline{y}$.

Although for both channel models a different likelihood function is obtained, a comparison of Equation 5.6 and Equation 5.9 reveals that their maximization is obtained by the minimization of the squared Euclidean distances of the receive sequence $\underline{y}$ to the incorrupted sequences ($\underline{x}$ in case of AWGN model 1 and $\overline{\underline{x}}$ in case of AWGN model 2).

Rayleigh Channel

In Chapter 3 various fading channel models were presented. They differ by the distribution of the fading (Rayleigh and multiple Rayleigh) as well as by the temporal correlation of the fading. Yet, as already mentioned in Section 3.3.2, if the amplitude fading $\mathbf{r}\left[k\right] = r\left[k\right]$ of a receive symbol $y\left[k\right]$ is known (perfect channel state information CSI), the Rayleigh channel is nothing but an AWGN channel with an effective signal-energy-to-noise-power-density ratio given by Equation 3.15. Hence, maximum likelihood decoding of a receive sequence $\underline{\mathbf{y}} = \underline{y}$, which has been passed through a Rayleigh channel with known amplitude fading $\underline{r}$, is choosing the valid channel symbol sequence which maximizes the likelihood function $f_{\underline{\mathbf{y}}|\overline{\underline{x}},\underline{r}}\left(\underline{y}\middle|\,\overline{\underline{x}}, \underline{r}\right)$. This is independent of the correlation among the amplitude fading sequence. Also, if the CSI is known, the AWGN is the only remaining noise and the likelihood function can also be expressed as the product of the

individual conditioned symbol PDFs

$$f_{\underline{\mathbf{y}}|\underline{\overline{x}},\underline{r}}\left(\underline{y}\middle|\,\underline{\overline{x}},\underline{r}\right)=\prod_{k=0}^{L-1}f_{\mathbf{y}|\overline{x},r}\left(y\left[k\right]\middle|\,\overline{x}\left[k\right],r\left[k\right]\right) \tag{5.10}$$

If again the natural logarithm is taken, the log-likelihood function for the Rayleigh channel with perfect CSI is obtained

$$ln\left(f_{\underline{\mathbf{y}}|\underline{\overline{x}},\underline{r}}\left(\underline{y}\middle|\,\underline{\overline{x}},\underline{r}\right)\right)=\sum_{k=0}^{L-1}ln\left(f_{\mathbf{y}|\overline{x},r}\left(y\left[k\right]\middle|\,\overline{x}\left[k\right],r\left[k\right]\right)\right). \tag{5.11}$$

In Section 3.3.2 two discrete Rayleigh channel models were presented. When Equation 3.14 is substituted into Equation 5.11, the log-likelihood function of **Rayleigh model 1** (see Figure 3.8 on Page 33) is obtained to

$$\begin{aligned} ln\left(f_{\underline{\mathbf{y}}|\underline{\overline{x}},\underline{r}}\left(\underline{y}\middle|\,\underline{\overline{x}},\underline{r}\right)\right) &= \textstyle\sum_{k=0}^{L-1}ln\left(\frac{\overline{E_S^{(Rx)}}}{\pi\cdot N_0}\cdot e^{-\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot\left|y[k]-r[k]\cdot\overline{x}[k]\right|^2}\right)\\ &= -\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot\textstyle\sum_{k=0}^{L-1}\left|y\left[k\right]-r\left[k\right]\cdot\overline{x}\left[k\right]\right|^2+L\cdot ln\left(\frac{\overline{E_S^{(Rx)}}}{\pi\cdot N_0}\right)\\ &= -K_5\cdot\textstyle\sum_{k=0}^{L-1}d_E^2\left(y\left[k\right],r\left[k\right]\cdot\overline{x}\left[k\right]\right)+K_6. \end{aligned}$$

Since $K_5>0$, ML decoding of a receive sequence $\underline{\mathbf{y}}=\underline{y}$ resulting from passing a normalized transmit sequence through Rayleigh channel model 1 becomes identical to the selection of the valid channel symbol sequence $\underline{x}$, which minimizes the squared Euclidean distance between the receive sequence $\underline{\mathbf{y}}=\underline{y}$ and the Rayleigh fading corrupted sequence

$$\underline{r}\cdot\underline{\overline{x}}=\left(r\left[0\right]\cdot\overline{x}\left[0\right],r\left[1\right]\cdot\overline{x}\left[1\right],\ldots,r\left[k\right]\cdot\overline{x}\left[k\right],\ldots,r\left[L-1\right]\cdot\overline{x}\left[L-1\right]\right)$$

without the additive noise:

$$\begin{aligned} d_E^2\left(\underline{r}\cdot\underline{\overline{x}},\underline{y}\right) &= \textstyle\sum_{k=0}^{L-1}\left|y\left[k\right]-r\left[k\right]\cdot\overline{x}\left[k\right]\right|^2\\ &= \textstyle\sum_{k=0}^{L-1}d_E^2\left(r\left[k\right]\cdot\overline{x}\left[k\right],y\left[k\right]\right). \end{aligned} \tag{5.12}$$

For the **second Rayleigh model**, presented in Figure 3.9 on Page 34, the substitution of Equa-

tion 3.16 in Equation 5.11 yields

$$\begin{aligned} ln\left(f_{\underline{\mathbf{y}}|\underline{\overline{x}},\underline{r}}\left(\underline{y}|\,\underline{\overline{x}},\underline{r}\right)\right) &= \sum_{k=0}^{L-1} ln\left(\frac{r^2[k]}{\pi}\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot e^{-\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot r^2[k]\cdot|y[k]-\overline{x}[k]|^2}\right) \\ &= -\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot\sum_{k=0}^{L-1} r^2\left[k\right]\cdot\left|y\left[k\right]-\overline{x}\left[k\right]\right|^2 + L\cdot ln\left(\frac{\overline{E_S^{(Rx)}}}{\pi\cdot N_0}\right)+\sum_{k=0}^{L-1} r^2\left[k\right] \\ &= -K_7\cdot\sum_{k=0}^{L-1} r^2\left[k\right]\cdot d_E^2\left(y\left[k\right],\overline{x}\left[k\right]\right)+K_8. \end{aligned}$$

Hence, with $K_7 > 0$ and both constants K_7 and K_8 independent of the valid channel sequence $\underline{\overline{x}}$, the maximum likelihood and the log-likelihood functions are maximized by the sequence $\underline{\overline{x}}$ which minimizes the weighted sum of the squared symbol Euclidean distances

$$\sum_{k=0}^{L-1} r^2\left[k\right]\cdot d_E^2\left(y\left[k\right],\overline{x}\left[k\right]\right). \tag{5.13}$$

Now a comparison of Equation 5.12 and 5.13 reveals that the sum, which needs to be minimized in order to maximize the likelihood and the log-likelihood functions, is dependent on the channel model! The reader should be aware that other valid implementations of a Rayleigh fading channel are possible and that for each implementation the log-likelihood function must be derived in order to obtain the correct sum composed of $y\left[k\right]$, $\overline{x}\left[k\right]$, and $r\left[k\right]$. This becomes especially relevant if the system has an automatic gain control (AGC).

Multiple Rayleigh Channels

The Multiple Rayleigh channel model, depicted in Figure 3.10 on Page 39, differs from Rayleigh model 2 only in the way the effective signal-energy-to-noise-power-density ratio $\frac{E_S}{N_0}\left[k\right]$ is generated. The CSI, however, consists of l amplitude vectors $\underline{r_0},\ldots\underline{r_{l-1}}$ with $\underline{r_j} = \left(r_j\left[0\right],\ldots,r_j\left[L-1\right]\right)$, $j = 0,\ldots,l-1$, one for each of the l main paths of the channel model, resulting in a single effective signal-energy-to-noise-power-density ratio vector

$$\underline{\frac{E_S}{N_0}} = \sum_{j=0}^{l-1}\overline{p_j}\cdot\underline{r_j}.$$

The conditioned PDF that a output $\mathbf{y} = y$ is observed under the constraint of the CSI and the transmitted symbol $\overline{x}$ is given by Equation 3.19. Then again, the additive white noise is the only remaining random variable and the maximum likelihood function is the product of the individual conditioned PDFs:

$$f_{\underline{\mathbf{y}}|\underline{\overline{x}},\underline{r_0},\ldots,\underline{r_{l-1}}}\left(\underline{y}|\,\underline{\overline{x}},\underline{r_0},\ldots,\underline{r_{l-1}}\right) = \prod_{k=0}^{L-1} f_{\mathbf{y}|\overline{x},r_0,\ldots,r_{l-1}}\left(y\left[k\right]|\,\overline{x}\left[k\right],r_0\left[k\right],\ldots,r_{l-1}\left[k\right]\right). \tag{5.14}$$

Hence, if in Equation 3.19 is substituted into Equation 5.14 the likelihood function of the presented Multiple Rayleigh channel model is obtained to

$$\begin{aligned} f_{\underline{\mathbf{y}}|\underline{\overline{x}},r_0,\ldots,r_{l-1}}\left(\underline{y}|\,\underline{\overline{x}},\underline{r_0},\ldots,\underline{r_{l-1}}\right) &= f_{\underline{\mathbf{y}}|\underline{\overline{x}},\underline{\frac{\mathbf{E}_S}{\mathbf{N}_0}}}\left(\underline{y}|\,\underline{\overline{x}},\underline{\frac{E_S}{N_0}}\right) \\ &= \tfrac{1}{\pi^L}\cdot\textstyle\prod_{k=0}^{L-1}\left[\frac{E_S}{N_0}[k]\right]\cdot e^{-\sum_{k=0}^{L-1}\frac{E_S}{N_0}[k]\cdot|y[k]-\overline{x}[k]|^2} \end{aligned} \tag{5.15}$$

with

$$\frac{E_S}{N_0}[k]=\sum_{j=0}^{l-1}\overline{p_j}\cdot\mathbf{r}_j^2[k]\cdot\frac{\overline{E_S^{(Tx)}}}{N_0}.$$

Correspondingly, the log-likelihood function is obtained to

$$\begin{aligned} ln\left(f_{\underline{\mathbf{y}}|\underline{\overline{x}},\underline{\frac{\mathbf{E}_S}{\mathbf{N}_0}}}\left(\underline{y}|\,\underline{\overline{x}},\underline{\frac{E_S}{N_0}}\right)\right) &= ln\left(\tfrac{1}{\pi^L}\cdot\textstyle\prod_{k=0}^{L-1}\left[\frac{E_S}{N_0}[k]\cdot e^{-\frac{E_S}{N_0}[k]\cdot|y[k]-\overline{x}[k]|^2}\right]\right) \\ &= -\textstyle\sum_{k=0}^{L-1}\frac{E_S}{N_0}[k]\cdot|y[k]-\overline{x}[k]|^2+ln\left(\tfrac{1}{\pi^L}\cdot\prod_{k=0}^{L-1}\frac{E_S}{N_0}[k]\right) \\ &= -\textstyle\sum_{k=0}^{L-1}\frac{E_S}{N_0}[k]\cdot d_E^2(y[k],\overline{x}[k])+K_9. \end{aligned}$$

Hence, the maximum of the likelihood and the log-likelihood function is obtained if the sum of weighted squared Euclidean distances

$$\sum_{k=0}^{L-1}\frac{E_S}{N_0}[k]\cdot d_E^2(y[k],\overline{x}[k])$$

is minimized. Again, this result is valid only for the presented implementation of a multiple Rayleigh channel.

5.1.2 Maximum Likelihood Decoding of Multiple Transmissions

So far, we treated only the maximum likelihood decoding of a single transmission. We now move to the more general case, that the information sequence $\underline{b}$ is encoded onto the code sequence $\underline{c}=g_c(\underline{b})$, which is itself mapped onto a transmission sequence $\underline{x^j}$ via a possibly variable one-to-one mapping g_m^j

$$\begin{aligned} \underline{x^j} &= g_m^j(\underline{c}) \\ &= g_m^j(g_c(\underline{b})), \end{aligned}$$

i.e. in our treatment we restrict ourself to the same code for the various transmission, but we allow for varying mappings of the channel symbol labels onto the channel symbols. The question which now arises is how a maximum likelihood decoding is performed if N mutually independent receive sequences $\underline{\overline{y}^j}$, $j=0,\ldots,N-1$ are available with the corresponding perfect CSIs,

obtained by passing the N transmit sequences $\underline{x^j}$ through one of the channel model presented in Chapter 3.

As in the previous section, we start with the AWGN channel model.

AWGN Channel

The log-likelihood function of **AWGN channel model 1** (Figure 3.2 on Page 26) was given in Equation 5.4. It is the PDF that the sequence $\underline{y}$ is received under the constrained that the cannel sequence $\underline{x}$ was sent. For the decoding process, however, the PDF that the sequence $\underline{y}$ is received under the constraint that the information sequence $\underline{b}$ was sent is of concern, since our goal is to determine the most likely information sequence. Yet, due to the one-to-one mapping g_c of the encoder and the one-to-one mapping g_m^j of the mapper the coresponding conditioned PDFs are identical, i.e.

$$f_{\underline{\mathbf{y}}|\underline{b}}\left(\underline{y^j}\middle|\,\underline{b}\right) = f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y^j}\middle|\,g_m^j\left(g_c\left(\underline{b}\right)\right)\right).$$

In this section, however, we use different mappings and accordingly can not use the PDFs conditioned on the transmit sequence. However, since all transmissions are based on the same code sequence we can use the PDF conditioned on the code sequence for maximum likelihood decoding. Then, the conditioned PDF that the j-th receive sequence $\underline{y^j}$ is received under the condition that the code sequence $\underline{c}$ was transmitted is

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}}}\left(\underline{y^j}\middle|\,\underline{c}\right) = f_{\underline{\mathbf{y}}|\underline{x}}\left(\underline{y^j}\middle|\,g_m^j\left(\underline{c}\right)\right).$$

Substituting Equation 5.4 yields for the conditioned PDF

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}}}\left(\underline{y^j}\middle|\,\underline{c}\right) = \frac{1}{\left[\pi\cdot N_0^j\right]^L}\cdot e^{-\frac{1}{N_0^j}\cdot\sum_{k=0}^{L-1}\left|y[k]-g_m^j(c[k])\right|^2}, \tag{5.16}$$

where N_0^j represents the spectral noise power density of the AWGN channel during the j-th transmission.

If N mutually independent receive sequences $\underline{y_0}, \underline{y_1}, \ldots, \underline{y_{N-1}}$ are available as a result of the transmissions of N channel sequences $g_m^0\left(\underline{c}\right), g_m^1\left(\underline{c}\right), \ldots, g_m^{N-1}\left(\underline{c}\right)$ over the AWGN channel model 1 with spectral power densities $N_0^0, N_0^1, \ldots, N_0^{N-1}$, the likelihood function

$$f_{\underline{\mathbf{y}^0},\underline{\mathbf{y}^1},\ldots,\underline{\mathbf{y}^{N-1}}|\underline{\mathbf{c}}}\left(\underline{y^0},\underline{y^1},\ldots,\underline{y^{N-1}}\middle|\,\underline{c}\right)$$

is the product of the N conditioned PDFs $f_{\underline{\mathbf{y}}|\underline{\mathbf{c}}}\left(\underline{y^0}\middle|\,\underline{c}\right), f_{\underline{\mathbf{y}}|\underline{\mathbf{c}}}\left(\underline{y^1}\middle|\,\underline{c}\right), \ldots, f_{\underline{\mathbf{y}}|\underline{\mathbf{c}}}\left(\underline{y^{N-1}}\middle|\,\underline{c}\right)$:

$$\begin{aligned} f_{\underline{\mathbf{y}}^0,\underline{\mathbf{y}}^1,\ldots,\underline{\mathbf{y}}^{N-1}|\underline{\mathbf{c}}}\left(\underline{y}^0,\underline{y}^1,\ldots,\underline{y}^{N-1}\middle|\,\underline{c}\right) &= \prod_{j=0}^{N-1} f_{\underline{\mathbf{y}}|\underline{\mathbf{c}}}\left(\underline{y}^j\middle|\,\underline{c}\right) \\ &= \frac{1}{\pi^{N\cdot L}}\cdot\frac{1}{\left[\prod_{j=0}^{N-1} N_0^j\right]^L}\cdot e^{-\sum_{j=0}^{N-1}\left[\frac{1}{N_0^j}\sum_{k=0}^{L-1}\left|y^j[k]-g_m^j(c[k])\right|^2\right]} \\ &= K_{10}\cdot e^{-\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot\sum_{k=0}^{L-1} d_E^2\left(y^j[k],g_m^j(c[k])\right)},\ K_{10}>0. \end{aligned}$$

Then, the log-likelihood function is obtained to

$$ln\left(f_{\underline{\mathbf{y}}^0,\underline{\mathbf{y}}^1,\ldots,\underline{\mathbf{y}}^{N-1}|\underline{\mathbf{c}}}\left(\underline{y}^0,\underline{y}^1,\ldots,\underline{y}^{N-1}\middle|\,\underline{c}\right)\right) = ln\left(K_{10}\right) - \sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot\sum_{k=0}^{L-1} d_E^2\left(y^j\left[k\right],g_m^j\left(c\left[k\right]\right)\right)$$

and ML decoding is equivalent to finding the valid code sequence $\underline{c}=\left(c\left[0\right],\ldots,c\left[L-1\right]\right)$ which minimizes

$$\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot\sum_{k=0}^{L-1} d_E^2\left(y^j\left[k\right],g_m^j\left(c\left[k\right]\right)\right). \tag{5.17}$$

Equation 5.17 reflects that squared distance contributions from more reliable transmissions (small N_0^j) are emphasized, whereas contributions from unreliable transmissions (large N_0^j) are down-weighted.

For **AWGN channel model 2**, the PDF of the receive sequence $\underline{y^j}$ under the condition that the code sequence $\underline{c}$ and hence the channel sequence $\overline{g}_m^j\left(c\left[k\right]\right)$ was sent is (compare Equation 5.7)

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}}}\left(\underline{y^j}\middle|\,\underline{c}\right) = \left[\frac{1}{\pi}\cdot\frac{E_S}{N_0}\middle|^j\right]^L\cdot e^{-\frac{E_S}{N_0}\middle|^j\cdot\sum_{k=0}^{L-1}\left|y^j[k]-\overline{g}_m^j(c[k])\right|^2},$$

where $\frac{E_S}{N_0}\Big|^j$ represents the signal energy to noise power density ratio of the j-th transmission and $\overline{g}_m^j$ the j-th mapping onto the normalized alphabet. Using the same argumentation as for model 1, the likelihood function can be obtained as product of the individual conditioned PDFs to

$$\begin{aligned} &f_{\underline{\mathbf{y}}^0,\underline{\mathbf{y}}^1,\ldots,\underline{\mathbf{y}}^{N-1}|\underline{\mathbf{c}}}\left(\underline{y}^0,\underline{y}^1,\ldots,\underline{y}^{N-1}\middle|\,\underline{c}\right) \\ &\quad= \prod_{j=0}^{N-1} f_{\underline{\mathbf{y}}|\underline{\mathbf{c}}}\left(\underline{y}^j\middle|\,\underline{c}\right) \\ &\quad= \frac{1}{\pi^{N\cdot L}}\cdot\left[\prod_{j=0}^{N-1}\frac{E_S}{N_0}\middle|^j\right]^L\cdot e^{-\sum_{j=0}^{N-1}\left[\frac{E_S}{N_0}\middle|^j\cdot\sum_{k=0}^{L-1}\left|y^j[k]-\overline{g}_m^j(c[k])\right|^2\right]} \\ &\quad= K_{11}\cdot e^{-\sum_{j=0}^{N-1}\frac{E_S}{N_0}\middle|^j\cdot\sum_{k=0}^{L-1} d_E^2\left(y^j[k],\overline{g}_m^j(c[k])\right)},\ K_{11}>0. \end{aligned}$$

and the log-likelihood function to

$$ln\left(f_{\underline{\mathbf{y}}^0,\underline{\mathbf{y}}^1,\ldots,\underline{\mathbf{y}}^{N-1}|\underline{\mathbf{c}}}\left(\underline{y}^0,\underline{y}^1,\ldots,\underline{y}^{N-1}|\,\underline{c}\right)\right) = ln\left(K_{11}\right) - \sum_{j=0}^{N-1}\left|\frac{E_S}{N_0}\right|^j \cdot \sum_{k=0}^{L-1} d_E^2\left(y^j\left[k\right],\overline{g}_m^j\left(c\left[k\right]\right)\right).$$

Accordingly, maximum likelihood decoding of N receive sequences obtained by sending a specific codeword with N normalized mappings through the AWGN channel model 2 is identical to finding the valid code sequence $\underline{c}$ which minimizes

$$\sum_{j=0}^{N-1}\left|\frac{E_S}{N_0}\right|^j \cdot \sum_{k=0}^{L-1} d_E^2\left(y^j\left[k\right],\overline{g}_m^j\left(c\left[k\right]\right)\right). \tag{5.18}$$

Again, this equations reflects the proper emphasis or down-weighting for reliable (large $\left|\frac{E_S}{N_0}\right|^j$) or unreliable (small $\left|\frac{E_S}{N_0}\right|^j$) transmissions, respectively.

In the preceeding section we saw that for the maximum likelihood decoding of a single receive sequence the actual AWGN model was irrelevant. Although both presented channel models resulted in different log-likelihood functions (Equations 5.3 and 5.5), they resulted in the same term, which needed to be minimized (Equations 5.6 and 5.9). If, on the other hand, several transmissions are received and if the channel statistics (N_0^j for model 1 and $\left|\frac{E_S}{N_0}\right|^j$ for model 2) are varying for these transmissions then the weighting factor become channel model dependent.

As a consequence, if a practical system makes use of several transmission with different SNR, it must be determined whether these different ratios are a result of different receive signal energies or different noise power densities, or both. The most general way to determine the correct weighting factors is to analyse the corresponding model and derive the log-likelihood function as presented in this section.

However, if the transmissions were passed through the AWGN channel model 2 with constant ratio $\frac{E_S}{N_0}$, maximum likelihood decoding becomes selecting the valid code sequence which minimizes

$$\sum_{j=0}^{N-1}\sum_{k=0}^{L-1} d_E^2\left(y^j\left[k\right],\overline{g}_m^j\left(c\left[k\right]\right)\right) \tag{5.19}$$

and no weighting is required. In fact, Equations 5.26 and 5.19 are identical.

Rayleigh Channel

In analogy with the AWGN channel model, if several transmissions $\underline{y}^j$ of a code sequence $\underline{c}$ with different normalized mappings $\overline{g}_c^j$ are received after being passed through a Rayleigh channel with arbitrary temporal fading but known CSI $\underline{r}^j$, the PDF of the individual receive sequences conditioned on the sent code sequence is related to the PDF of the receive sequences conditioned on the sent channel sequence as follows (one-to-one mapping of the mapper)

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y^j}\middle|\,\underline{c},\underline{r^j}\right)=f_{\underline{\mathbf{y}}|\underline{\overline{x}},\underline{r}}\left(\underline{y^j}\middle|\,g_m^j\left(\underline{c}\right),\underline{r^j}\right).$$

For **Rayleigh model 1** (see Figure 3.8 on Page 33), with Equation 3.14, Equation 5.10, and the above notation, the conditioned PDF for the j-th transmission can be obtained to

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y^j}\middle|\,\underline{c},\underline{r^j}\right)=\left(\frac{1}{\pi}\cdot\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\right)^L\cdot e^{-\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\cdot\sum_{k=0}^{L-1}\left|y^j[k]-r^j[k]\cdot\overline{g}_c^j(c[k])\right|^2}.$$

Like in the AWGN case, we allow for different (average) signal-energy-to-noise-power-density ratios $\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j$ during the individual transmissions.

Now, if N mutually independent receive sequences $\underline{y_0},\underline{y_1},\ldots,\underline{y_{N-1}}$ are available as a result of the transmissions of N channel sequences $\overline{g}_m^0\left(\underline{c}\right),\overline{g}_m^1\left(\underline{c}\right),\ldots,\overline{g}_m^{N-1}\left(\underline{c}\right)$ over the Rayleigh channel model 1 with average ratios densities $\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^0,\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^1,\ldots,\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^{N-1}$, the likelihood function

$$f_{\underline{\mathbf{y}^0},\underline{\mathbf{y}^1},\ldots,\underline{\mathbf{y}^{N-1}}|\underline{\mathbf{c}},\underline{r^0},\underline{r^1},\ldots,\underline{r^{N-1}}}\left(\underline{y^0},\underline{y^1},\ldots,\underline{y^{N-1}}\middle|\,\underline{c},\underline{r^0},\underline{r^1},\ldots,\underline{r^{N-1}}\right)$$

is the product of the N conditioned PDFs

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y^0}\middle|\,\underline{c},\underline{r^0}\right),f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y^1}\middle|\,\underline{c},\underline{r^1}\right),\ldots,f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y^{N-1}}\middle|\,\underline{c},\underline{r^{N-1}}\right).$$

Hence, the likelihood function is

$$\begin{aligned}
&f_{\underline{\mathbf{y}^0},\underline{\mathbf{y}^1},\ldots,\underline{\mathbf{y}^{N-1}}|\underline{\mathbf{c}},\underline{r^0},\underline{r^1},\ldots,\underline{r^{N-1}}}\left(\underline{y^0},\underline{y^1},\ldots,\underline{y^{N-1}}\middle|\,\underline{c},\underline{r^0},\underline{r^1},\ldots,\underline{r^{N-1}}\right)\\
&\quad=\prod_{j=0}^{N-1}f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y^j}\middle|\,\underline{c},\underline{r^j}\right)\\
&\quad=\tfrac{1}{\pi^{N\cdot L}}\cdot\left(\prod_{j=0}^{N-1}\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\right)^L\cdot e^{-\sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\cdot\sum_{k=0}^{L-1}\left|y^j[k]-r^j[k]\cdot\overline{g}_c^j(c[k])\right|^2\right]}\\
&\quad=K_{12}\cdot e^{-\sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\cdot\sum_{kj=0}^{L-1}d_E^2\left(y^j[k],r^j[k]\cdot\overline{g}_c^j(c[k])\right)\right]}
\end{aligned}$$

and the log-likelihood function

$$ln\left(f_{\underline{\mathbf{y}}^0,\underline{\mathbf{y}}^1,\ldots,\underline{\mathbf{y}}^{N-1}|\underline{\mathbf{c}},\underline{r}^0,\underline{r}^1,\ldots,\underline{r}^{N-1}}\left(\underline{y}^0,\underline{y}^1,\ldots,\underline{y}^{N-1}\middle|\,\underline{c},\underline{r}^0,\underline{r}^1,\ldots,\underline{r}^{N-1}\right)\right)$$

$$= ln\left(K_{12}\right) - \sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j \cdot \sum_{k=0}^{L-1} d_E^2\left(y^j\left[k\right], r^j\left[k\right]\cdot\overline{g}_c^j\left(c\left[k\right]\right)\right)\right].$$

Both likelihood functions reveal, that maximum likelihood decoding is identical to finding the valid code sequence $\underline{c} = \left(c\left[0\right],\ldots,c\left[L-1\right]\right)$ which minimizes

$$\sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j \cdot \sum_{k=0}^{L-1} d_E^2\left(y^j\left[k\right], r^j\left[k\right]\cdot\overline{g}_c^j\left(c\left[k\right]\right)\right)\right]. \tag{5.20}$$

Equation 5.20 states that for ML decoding of several channel words derived from the same codeword and passed through Rayleigh model 2, the squared Euclidean distances from the received channel symbol to the fading corrupted channel symbol must be computed and weighted by the corresponding average signal-energy-to-noise-power-density ratio.

For the **Rayleigh channel model 2** (Figure 3.9 on Page 34), the conditioned PDF $f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y}^j\middle|\,\underline{c},\underline{r}^j\right)$ becomes (Equation 3.16 and Equation 5.10)

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y}^j\middle|\,\underline{c},\underline{r}^j\right) = \left(\frac{1}{\pi}\cdot\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\right)^L \cdot \prod_{k=0}^{L-1}\left(r^j\left[k\right]\right)^2 \cdot e^{-\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\cdot\sum_{k=0}^{L-1}\left(r^j[k]\right)^2\cdot\left|y^j[k]-\overline{g}_c^j(c[k])\right|^2}.$$

Again, the likelihood function can be expressed as product of the individual conditioned PDFs

$$f_{\underline{\mathbf{y}}^0,\underline{\mathbf{y}}^1,\ldots,\underline{\mathbf{y}}^{N-1}|\underline{\mathbf{c}},\underline{r}^0,\underline{r}^1,\ldots,\underline{r}^{N-1}}\left(\underline{y}^0,\underline{y}^1,\ldots,\underline{y}^{N-1}\middle|\,\underline{c},\underline{r}^0,\underline{r}^1,\ldots,\underline{r}^{N-1}\right)$$

$$= \prod_{j=0}^{N-1} f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r}}\left(\underline{y}^j\middle|\,\underline{c},\underline{r}^j\right)$$

$$= \frac{1}{\pi^{N\cdot L}}\cdot\left(\prod_{j=0}^{N-1}\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\right)^L \cdot \prod_{j=0}^{N-1}\prod_{k=0}^{L-1} r^j\left[k\right]^2 \cdot e^{-\sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\cdot\sum_{k=0}^{L-1} r^j[k]^2\cdot\left|y^j[k]-\overline{g}_c^j(c[k])\right|^2\right]}$$

$$= K_{13}\cdot e^{-\sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j\cdot\sum_{k=0}^{L-1}\left(r^j[k]\right)^2\cdot d_E^2\left(y^j[k],\overline{g}_c^j(c[k])\right)\right]}$$

and for the corresponding log-likelihood function follows

$$ln\left(f_{\underline{\mathbf{y}}^0,\underline{\mathbf{y}}^1,\ldots,\underline{\mathbf{y}}^{N-1}|\underline{\mathbf{c}},\underline{r}^0,\underline{r}^1,\ldots,\underline{r}^{N-1}}\left(\underline{y}^0,\underline{y}^1,\ldots,\underline{y}^{N-1}\middle|\,\underline{c},\underline{r}^0,\underline{r}^1,\ldots,\underline{r}^{N-1}\right)\right)$$

$$= ln\left(K_{13}\right) - \sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j \cdot \sum_{k=0}^{L-1}\left(r^j\left[k\right]\right)^2 \cdot d_E^2\left(y^j\left[k\right],\overline{g}_c^j\left(c\left[k\right]\right)\right)\right].$$

Hence, ML decoding is identical to finding the valid code sequence $\underline{c} = \left(c\left[0\right],\ldots,c\left[L-1\right]\right)$ which minimizes

$$\sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j \cdot \sum_{k=0}^{L-1}\left(r^j\left[k\right]\right)^2 \cdot d_E^2\left(y^j\left[k\right],\overline{g}_c^j\left(c\left[k\right]\right)\right)\right]. \tag{5.21}$$

Therefore, for Rayleigh model 2, the squared Euclidean distance between the receive symbols and the transmitted symbols must be computed and weighted by the average receive signal-energy-to-noise-power-density ratio times the squared amplitude fading.

Multiple Rayleigh Channels

In Section 5.1.1, the likelihood function for a single transmission passed through the Multiple Rayleigh channel (see Figure 3.10 on Page 39) was derived under the condition that the CSI (l amplitude fading sequences $\underline{r_0},\ldots,\underline{r_{l-1}}$ or the resulting signal-energy-to-noise-power-density ratio sequence $\underline{\frac{E_S}{N_0}}$) are known. If N transmissions $\underline{y}^j$ of a code sequence $\underline{c}$ with different normalized mappings $\overline{g}_c^j$ are received after being passed through the Multiple Rayleigh channel, the complete CSI is characterized by a set R_l^N of $N\cdot l$ amplitude fading sequences $\underline{r_i^j}$, $j=0,\ldots,N-1$; $i=0,\ldots,l-1$ of length L or N signal-energy-to-noise-power-density ratios $\underline{\frac{E_S}{N_0}}^j$ of length L.

As previously, the PDF of a single receive sequence conditioned on the sent code sequence is related to the PDF of the receive sequences conditioned on the sent channel sequence as follows (one-to-one mapping of the mapper)

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r_0},\ldots,\underline{r_{l-1}}}\left(\underline{y^j}\middle|\,\underline{c},\underline{r_0^j},\ldots,\underline{r_{l-1}^j}\right) = f_{\underline{\mathbf{y}}|\underline{\overline{\mathbf{x}}},\underline{r_0},\ldots,\underline{r_{l-1}}}\left(\underline{y^j}\middle|\,g_m^j\left(\underline{c}\right),\underline{r_0^j},\ldots,\underline{r_{l-1}^j}\right)$$

or conditioned on the effective ratio

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{\frac{E_S}{N_0}}}\left(\underline{y^j}\middle|\,\underline{c},\underline{\frac{E_S}{N_0}}^j\right) = f_{\underline{\mathbf{y}}|\underline{\overline{\mathbf{x}}},\underline{\frac{E_S}{N_0}}}\left(\underline{y^j}\middle|\,g_m^j\left(\underline{c}\right),\underline{\frac{E_S}{N_0}}^j\right).$$

Substituting Equation 5.15 yields

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{c}},\underline{r_0},\ldots,\underline{r_{l-1}}}\left(\underline{y^j}\middle|\,\underline{c},\underline{r_0^j},\ldots,\underline{r_{l-1}^j}\right) = \frac{1}{\pi^L}\cdot\prod_{k=0}^{L-1}\left[\frac{E_S}{N_0}^j\left[k\right]\right]\cdot e^{-\sum_{k=0}^{L-1}\frac{E_S}{N_0}^j[k]\cdot\left|y^j[k]-g_m^j(\underline{c}[k])\right|^2}$$

with $\frac{E_S}{N_0}^j\left[k\right]$ being the effective ratio of the j-th transmission on symbol position k, i.e.

$$\frac{E_S}{N_0}^j [k] = \sum_{i=0}^{l-1} \overline{p_i}^j \cdot \mathbf{r}_i^j [k]^2 \cdot \frac{\overline{E_S^{(Rx)}}^j}{N_0} . \tag{5.22}$$

Like in the AWGN and the Rayleigh models, we allow for different average receive ratios $\frac{\overline{E_S^{(Rx)}}^j}{N_0}$. In addition, the channel profile $\left\{\overline{p_0}^j, \ldots, \overline{p_{l-1}}^j\right\}$ is also assumed to be variable for the various transmissions.

Now, if N mutually independent receive sequences $\underline{y^0}, \underline{y^1}, \ldots, \underline{y^{N-1}}$ are available as a result of the transmissions of N channel sequences $\overline{g}_m^0(\underline{c}), \overline{g}_m^1(\underline{c}), \ldots, \overline{g}_m^{N-1}(\underline{c})$ over the Multiple Rayleigh channel each with with average receive ratios $\left.\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^0, \ldots, \left.\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^{N-1}$ and normalized channel profiles $\left\{\overline{p_0}^0, \ldots, \overline{p_{l-1}}^0\right\}, \ldots, \left\{\overline{p_0}^{N-1}, \ldots, \overline{p_{l-1}}^{N-1}\right\}$, the likelihood function

$$f_{\underline{\mathbf{y}^0}, \underline{\mathbf{y}^1}, \ldots, \underline{\mathbf{y}^{N-1}} | \underline{\mathbf{c}}, \mathbf{R}_l^N} \left(\underline{y^0}, \underline{y^1}, \ldots, \underline{y^{N-1}} \middle| \underline{c}, R_l^N\right),$$

with $\mathbf{R}_l^N$ denoting the set of CSI, is the product of the N conditioned PDFs

$$f_{\underline{\mathbf{y}} | \underline{\mathbf{c}}, \underline{r_1}, \ldots, \underline{r_l}} \left(\underline{y^0} \middle| \underline{c}, \underline{r_1^0}, \ldots, \underline{r_l^0}\right), \ldots, f_{\underline{\mathbf{y}} | \underline{\mathbf{c}}, \underline{r_1}, \ldots, \underline{r_l}} \left(\underline{y^{N-1}} \middle| \underline{c}, \underline{r_1^{N-1}}, \ldots, \underline{r_l^{N-1}}\right).$$

Hence, the likelihood function is

$$\begin{aligned}
&f_{\underline{\mathbf{y}^0}, \underline{\mathbf{y}^1}, \ldots, \underline{\mathbf{y}^{N-1}} | \underline{\mathbf{c}}, \mathbf{R}_l^N} \left(\underline{y^0}, \underline{y^1}, \ldots, \underline{y^{N-1}} \middle| \underline{c}, R_l^N\right) \\
&\quad = \prod_{j=0}^{N-1} f_{\underline{\mathbf{y}} | \underline{\mathbf{c}}, \underline{r_0}, \ldots, \underline{r_{l-1}}} \left(\underline{y^j} \middle| \underline{c}, \underline{r_0^j}, \ldots, \underline{r_{l-1}^j}\right) \\
&\quad = \frac{1}{\pi^{N \cdot L}} \cdot \prod_{j=0}^{N-1} \prod_{k=0}^{L-1} \left[\left.\frac{E_S}{N_0}\right|^j [k]\right] \cdot e^{-\sum_{j=0}^{N-1} \sum_{k=0}^{L-1} \left[\left.\frac{E_S}{N_0}\right|^j [k] \cdot \left|y^j[k] - \overline{g}_c^j(c[k])\right|^2\right]} \\
&\quad = K_{14} \cdot e^{-\sum_{j=0}^{N-1} \sum_{k=0}^{L-1} \left[\left.\frac{E_S}{N_0}\right|^j [k] \cdot d_E^2\left(y^j[k], \overline{g}_c^j(c[k])\right)\right]}
\end{aligned}$$

with $\frac{E_S}{N_0}^j [k]$ given by Equation 5.22. The log-likelihood function is obtained to

$$\begin{aligned}
&ln\left(f_{\underline{\mathbf{y}^0}, \underline{\mathbf{y}^1}, \ldots, \underline{\mathbf{y}^{N-1}} | \underline{\mathbf{c}}, \mathbf{R}_l^N} \left(\underline{y^0}, \underline{y^1}, \ldots, \underline{y^{N-1}} \middle| \underline{c}, R_l^N\right)\right) \\
&\quad = ln(K_{14}) - \sum_{j=0}^{N-1} \sum_{k=0}^{L-1} \left[\left.\frac{E_S}{N_0}\right|^j [k] \cdot d_E^2\left(y^j[k], \overline{g}_c^j(c[k])\right)\right].
\end{aligned}$$

Therefore, ML decoding is identical to finding the valid code sequence $\underline{c} = (c[0], \ldots, c[L-1])$ which minimizes

$$\sum_{j=0}^{N-1} \sum_{k=0}^{L-1} \left[\left.\frac{E_S}{N_0}\right|^j [k] \cdot d_E^2\left(y^j[k], \overline{g}_c^j(c[k])\right)\right], \tag{5.23}$$

with $\frac{E_S}{N_0}^j[k]$ given by Equation 5.22.

5.1.3 Maximum Likelihood Distance Combining

In Section 5.1.1, the likelihood and the log-likelihood functions for a single receiveword, which has been received after a channelword has been passed through one of the channels, introduced in Chapter 3, has been derived. As discussed, for ML decoding, the valid codeword (or corresponding channel word), which maximizes these likelihood functions, must be found. In the field of coding theory, efficient techniques for locating these codewords have been found, at least for some codes. Such an example is the Viterbi-algorithm [Pro95] for the class of convolutional codes.

This algorithm is based on the fact that all code sequences can be represented by trellis with 2^v states, with v representing the total number of binary memory elements. All possible code sequences at a certain time step will end at one of these states, and all sequences have associated a certain log-likelihood value. Due to the multiplicity of the likelihood function (or the additivity of the log-likelihood function), from all the code sequences which end at a certain state, only the sequence with the largest likelihood function is of interest, since it will have a higher likelihood value also for further time steps. Hence, at a certain time step only 2^v so-called survivers have to be traced for the the future, exactly one for each state. Therefore, at each time step and each state, a decision for one surviver has to be made, based on the previous survivors and the possible new increments of the likelihood function, the so-called path metrics. If, for example, the Rayleigh channel model 2 is used, the path metrics at time step k are (compare Equation 5.13)

$$r^2[k] \cdot d_E^2\left(y[k], \overline{x}[k]\right).$$

Now, if several receivewords from different convolutional codes have to be ML combined, no simple trellis representation of the combined code structure exists in general and the Viterbi algorithm does not apply. If, however, different channel word are derived via varying mapper from the *same* codeword, all channel words can be represented by a simple trellis and again the Viterbi algorithm can be used. In this case, combined path metrics need to calculated and used for the survivor decision.

The question, which in this case arises, is how the path metrics have to be combined for maximum likelihood decoding. Therefore, in this section, these combined path metrics for all the discussed channel models are derived from the results of the preceeding section and their memory requirements are evaluated, i.e. the corresponding normalized receiver memory cost function (see Section 4.6.2) is determined.

AWGN Channel

For maximum likelihood decoding of multiple receive words, which have been passes through the **AWGN channel model 1**, Equation 5.17 is required to be minimized. Reordering the sum

j	$g_m^j\left(c\left[2\right]\right)$	N_0^j	$y^j\left[2\right]$	path metric
0	$1+0\cdot j$	0.5	$1.2+0.3\cdot j$	0.26
1	$-1+0\cdot j$	0.6	$-0.7-0.4\cdot j$	0.42
2	$0+j$	0.4	$0.1+1.1\cdot j$	0.05
combined				0.73

Table 5.1: Example of ML Combining for AWGN Channel Model 1.

yields

$$\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot\sum_{k=0}^{L-1}\cdot d_E^2\left(y^j\left[k\right],g_m^j\left(c\left[k\right]\right)\right)=\sum_{k=0}^{L-1}\left[\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot d_E^2\left(y^j\left[k\right],g_m^j\left(c\left[k\right]\right)\right)\right]$$

and after N transmissions, the combined path metrics for a codesymbol $c\left[k\right]$ is identical to the sum

$$\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot d_E^2\left(y^j\left[k\right],g_m^j\left(c\left[k\right]\right)\right). \tag{5.24}$$

Equation 5.24 states that for the AWGN channel model 1 at a certain time step k the squared Euclidean distances from the receive symbols $y^j\left[k\right]$ to the corresponding channel symbols $g_m^j\left(c\left[k\right]\right)$ have to be computed and divided by the noise-power-density N_0^j of the corresponding transmission. This division provides the proper likelihood scaling in order to account for more reliable receive sequences due to a small noise power density.

Table 5.1 depicts an example for a certain time step $k=2$. The noise-power-densities for the first transmission and 2 retransmissions, as well as the transmit channel symbols, and the receive symbols are shown. In the last column, the resulting individual path metrics and the ML combined path metrics according to Equation 5.24 are listed.

For the special case of a constant mapping g_m, Equation 5.24 simplifies to

$$\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot d_E^2\left(y^j\left[k\right],\overline{g}_m\left(c\left[k\right]\right)\right)=\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot d_E^2\left(y^j\left[k\right],x\left[k\right]\right) \tag{5.25}$$

and if, in addition, the channel has a constant spectral noise power density N_0 for all transmissions, maximum likelihood decoding is achieved if the code sequence is selected which minimizes

$$\sum_{k=0}^{L-1}\sum_{j=0}^{N-1}d_E^2\left(y^j\left[k\right],g_m^j\left(c\left[k\right]\right)\right), \tag{5.26}$$

that is the unweighted squared Euclidean sums.

For maximum likelihood decoding of multiple receive word, which have been passed through

AWGN channel model 2, Equation 5.18 needs to be minimized. Again, reordering the sum yields

$$\sum_{j=0}^{N-1} \left|\frac{E_S}{N_0}\right|^j \cdot \sum_{k=0}^{L-1} d_E^2\left(y^j[k], \overline{g}_m^j(c[k])\right) = \sum_{k=0}^{L-1}\sum_{j=0}^{N-1}\left[\left|\frac{E_S}{N_0}\right|^j \cdot d_E^2\left(y^j[k], \overline{g}_m^j(c[k])\right)\right]$$

Hence, after after N transmissions, the combined path metrics for a codesymbol $c[k]$ is identical to the sum

$$\sum_{j=0}^{N-1}\left[\left|\frac{E_S}{N_0}\right|^j \cdot d_E^2\left(y^j[k], \overline{g}_m^j(c[k])\right)\right]. \tag{5.27}$$

Equation 5.18 states that for maximum likelihood combining of multiple receive sequences passed through AWGN channel model 2, the squared Euclidean distances from the actual receive symbols $y^j[k]$ to the receive symbols without additive noise ($\overline{g}_m^j(c[k])$ in AWGN channel model 2) must be computed and multiplicatively weighted by the corresponding signal-energy-to-noise-power-density ratio $\left|\frac{E_S}{N_0}\right|^j$, before being added to the running sum. Again, this multiplication provides the correct likelihood scaling for more reliable receive sequences due to a higher SNR.

For the special case, that a constant normalized mapping $\overline{g}_m$ is used, Equation 5.27 simplifies to

$$\sum_{j=0}^{N-1}\left|\frac{E_S}{N_0}\right|^j \cdot d_E^2\left(y^j[k], \overline{g}_m(c[k])\right) = \sum_{j=0}^{N-1}\left|\frac{E_S}{N_0}\right|^j \cdot d_E^2\left(y^j[k], \overline{x}[k]\right) \tag{5.28}$$

and if, in addition, the SNR is constant, ML combining is simply the addition of the squared distances

$$\sum_{j=0}^{N-1} d_E^2\left(y^j[k], \overline{x}[k]\right).$$

Rayleigh Channel

For maximum likelihood decoding of multiple receive words, which have been passes through the **Rayleigh channel model 1**, Equation 5.20 is required to be minimized. Reordering the sum yields

$$\sum_{k=0}^{L-1}\sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j \cdot d_E^2\left(y^j[k], r^j[k]\cdot\overline{g}_c^j(c[k])\right)\right]$$

and after N transmissions, the combined path metrics for a codesymbol $c[k]$ is identical to the sum

$$\sum_{j=0}^{N-1}\left[\left|\frac{\overline{E_S^{(Rx)}}}{N_0}\right|^j \cdot d_E^2\left(y^j[k], r^j[k]\cdot\overline{g}_c^j(c[k])\right)\right]. \tag{5.29}$$

Equation 5.29 states that for maximum likelihood combining of multiple receive sequences passed through Rayleigh channel model 1, the squared Euclidean distances from the actual receive symbols $y^j[k]$ to the receive symbols without additive noise ($r^j[k]\cdot\overline{g}_m^j(c[k])$ in Rayleigh

channel model 1) must be computed and multiplicatively weighted by the corresponding signal-energy-to-noise-power-density ratio $\left|\overline{\frac{E_S^{(Rx)}}{N_0}}\right|^j$ before being added to the running sum.

If a constant mapper $\overline{g}_c\left(c\left[k\right]\right)$ is used for all transmissions, ML combining for Rayleigh channel model 1 can be simplified to

$$\sum_{j=0}^{N-1}\left[\left|\overline{\frac{E_S^{(Rx)}}{N_0}}\right|^j \cdot d_E^2\left(y^j\left[k\right], r^j\left[k\right] \cdot \overline{g}_c\left(c\left[k\right]\right)\right)\right] = \sum_{j=0}^{N-1}\left[\left|\overline{\frac{E_S^{(Rx)}}{N_0}}\right|^j \cdot d_E^2\left(y^j\left[k\right], r^j\left[k\right] \cdot \overline{x}\left[k\right]\right)\right], \quad (5.30)$$

and if, moreover, the average signal-energy-to-noise-power-density ratio is identical for all transmissions, it can be taken in front of the summation and is not relevant for the minimization. Then, for maximum likelihood combining is identical to

$$\sum_{j=0}^{N-1} d_E^2\left(y^j\left[k\right], r^j\left[k\right] \cdot \overline{g}_c\left(c\left[k\right]\right)\right).$$

For maximum likelihood decoding of multiple receive words, which have been passes through the **Rayleigh channel model** 2, Equation 5.21 is required to be minimized. Reordering the sum yields

$$\sum_{k=0}^{L-1}\sum_{j=0}^{N-1}\left[\left|\overline{\frac{E_S^{(Rx)}}{N_0}}\right|^j \cdot \left(r^j\left[k\right]\right)^2 \cdot d_E^2\left(y^j\left[k\right], \overline{g}_c^j\left(c\left[k\right]\right)\right)\right]$$

and after N transmissions, the combined path metrics for a codesymbol $c\left[k\right]$ is identical to the sum

$$\sum_{j=0}^{N-1}\left[\left|\overline{\frac{E_S^{(Rx)}}{N_0}}\right|^j \cdot \left(r^j\left[k\right]\right)^2 \cdot d_E^2\left(y^j\left[k\right], \overline{g}_c^j\left(c\left[k\right]\right)\right)\right]. \quad (5.31)$$

Again, Equation 5.31 states that for maximum likelihood combining of multiple receive sequences passed through Rayleigh channel model 2, the squared Euclidean distances from the actual receive symbols $y^j\left[k\right]$ to the receive symbols without additive noise ($\overline{g}_m^j\left(c\left[k\right]\right)$ in Rayleigh channel model 2) must be computed and multiplicatively weighted by the corresponding signal-energy-to-noise-power-density ratio $\left|\overline{\frac{E_S^{(Rx)}}{N_0}}\right|^j$ times the squared amplitude attenuation $\left(r^j\left[k\right]\right)^2$ before being added to the running sum.

If a constant mapper $\overline{g}_c\left(c\left[k\right]\right)$ is used for all transmissions, ML combining for Rayleigh channel model 2 can be simplified to

$$\begin{aligned}&\textstyle\sum_{j=0}^{N-1}\left[\left|\overline{\frac{E_S^{(Rx)}}{N_0}}\right|^j \cdot \left(r^j\left[k\right]\right)^2 \cdot d_E^2\left(y^j\left[k\right], \overline{g}_c^j\left(c\left[k\right]\right)\right)\right] \\ &\quad = \textstyle\sum_{j=0}^{N-1}\left[\left|\overline{\frac{E_S^{(Rx)}}{N_0}}\right|^j \cdot \left(r^j\left[k\right]\right)^2 \cdot d_E^2\left(y^j\left[k\right], \overline{x}\left[k\right]\right)\right],\end{aligned} \quad (5.32)$$

and if, moreover, the average signal-energy-to-noise-power-density ratio is identical for all transmissions, it can be taken in front of the summation and is not relevant for the minimization. Then, for maximum likelihood combining is identical to

$$\sum_{j=0}^{N-1} \left(r^j\left[k\right]\right)^2 \cdot d_E^2\left(y^j\left[k\right], \overline{g}_c^j\left(c\left[k\right]\right)\right).$$

Multiple Rayleigh Channel

For maximum likelihood decoding of multiple receive words, which have been passed through the Multiple Rayleigh channel model, Equation 5.23 is required to be minimized. Reordering the sum yields

$$\sum_{j=0}^{N-1}\sum_{k=0}^{L-1}\left[\left|\frac{E_S}{N_0}\right|^j\left[k\right] \cdot d_E^2\left(y^j\left[k\right], \overline{g}_c^j\left(c\left[k\right]\right)\right)\right]$$

and after N transmissions, the combined path metrics for a codesymbol $c\left[k\right]$ is identical to the sum

$$\sum_{k=0}^{L-1}\left[\left|\frac{E_S}{N_0}\right|^j\left[k\right] \cdot d_E^2\left(y^j\left[k\right], \overline{g}_c^j\left(c\left[k\right]\right)\right)\right], \tag{5.33}$$

with $\left|\frac{E_S}{N_0}\right|^j\left[k\right]$ given by Equation 5.22.

Equation 5.33 states that for maximum likelihood combining of multiple receive sequences passed through the Multiple Rayleigh channel model the squared Euclidean distances from the actual receive symbols $y^j\left[k\right]$ to the receive symbols without additive noise ($\overline{g}_m^j\left(c\left[k\right]\right)$ in the presented model) must be computed and multiplicatively weighted by the corresponding effective signal-energy-to-noise-power-density ratio $\left|\frac{E_S}{N_0}\right|^j\left[k\right]$ before being added to the running sum.

If a constant mapper $\overline{g}_c\left(c\left[k\right]\right)$ is used for all transmissions, ML combining for Rayleigh channel model 2 can be simplified to

$$\sum_{k=0}^{L-1}\left[\left|\frac{E_S}{N_0}\right|^j\left[k\right] \cdot d_E^2\left(y^j\left[k\right], \overline{g}_c\left(c\left[k\right]\right)\right)\right] = \sum_{k=0}^{L-1}\left[\left|\frac{E_S}{N_0}\right|^j\left[k\right] \cdot d_E^2\left(y^j\left[k\right], \overline{x}\left[k\right]\right)\right] \tag{5.34}$$

Memory Requirements

Throughout this section, the equations for maximum likelihood combining for the various channel models were derived. All equations consisted of the summation over weighted squared distances. In each time step, there are in general M possible transmit channel symbols (M is the size of the modulation alphabet) and, hence, M possible squared and weighted sums. Therefore, after the mother code rate is reached, $M \cdot L$ weighted squared distances, i.e. softdecision values or their integer quantizations, have to be stored. All further retransmitted symbols are ML combined according to the appropriate equation (Equations 5.24, 5.27, 5.29, 5.31, or 5.33). Hence, the

maximum receiver memory per transmitted information word is given by

$$M_p^{Rx}\big|_{max} = L \cdot M\,. \tag{5.35}$$

This maximum receiver memory requirement is also required if a constant mapper is used and if ML distance combining according to Equations 5.25, 5.28, 5.30, 5.32, and 5.34 is performed. In Section 5.1.4, however, a less memory extensive ML combining method for this special case is presented.

5.1.4 Maximum-Ratio-Combining

Maximum-Ratio-Combining (MRC) was already introduced in Section 3.4. However, only the effective result of MRC was stated:

> A maximum-ratio-diversity combining scheme combines N independent **replicas** of a signals received with the N individual signal-energy-to-noise-power-density ratios $\left.\frac{E_S^{(Rx)}}{N_0}\right|_j$, $j = 1,\ldots,N$ to the same signal with an effective signal-energy-to-noise-power-density ratio of
>
> $$\left.\frac{E_S^{(Rx)}}{N_0}\right|_{MRC} = \sum_{j=1}^{N} \left.\frac{E_S^{(Rx)}}{N_0}\right|_j \tag{5.36}$$

Based on this result the Multiple Rayleigh channel model, depicted in Figure 3.10 on Page 39, was derived.

In the following, it will be presented how MRC is accomplished and it will be shown that it is a form of maximum likelihood combining, i.e. that maximum likelihood decoding of a sequence $\underline{\mathbf{y}_{MRC}}$ resulting from maximum ratio combining of the sequences $\underline{\mathbf{y}^0},\ldots,\underline{\mathbf{y}^{N-1}}$ yields the same result as a direct maximum likelihood decoding of the individual sequences $\underline{\mathbf{y}^0},\ldots,\underline{\mathbf{y}^{N-1}}$.

Maximum Ratio Combining Scheme

Brennan investigated in his 1959 paper the general problem of combining copies of the same analog signal under the assumptions that the copies are

- independent
- corrupted by additive white noise
- attenuated by fading which is constant throughout the signal duration.

Among the three systems he presented, the so-called maximum ratio diversity system showed the best performance. Due to the linear operations of which MRC is composed of, Brennan's analysis for analog signals can be easily translated to digital symbols. Then, MRC combines

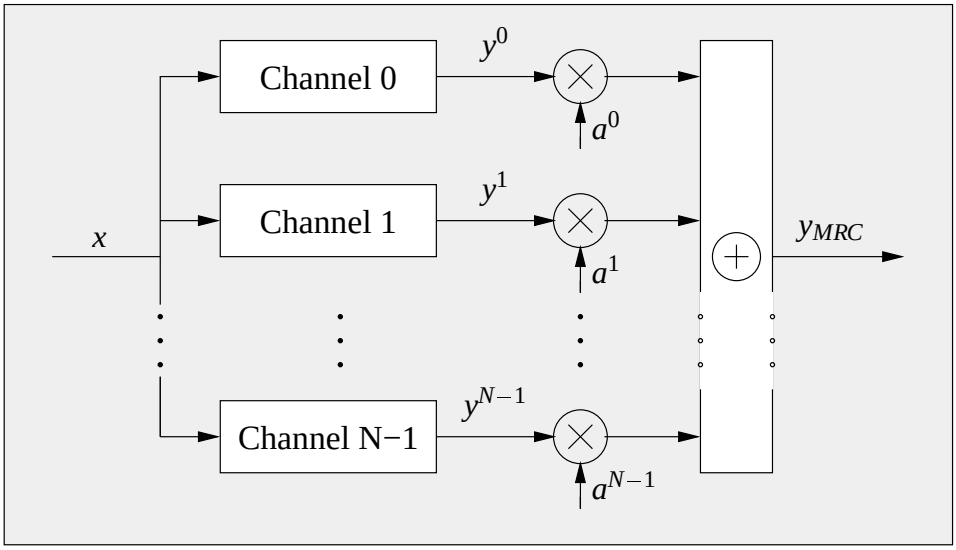

Figure 5.2: Block Diagram of Maximum Ratio Combining

	AWGN 1	AWGN 2	Rayleigh 1	Rayleigh 2
$E\{\mathbf{y}^j\}$	$\sqrt{E_S}$	1	r^j	1
$VAR\{\mathbf{y}^j\}$	$\frac{N_0^j}{2}$	$\left[2 \cdot \left\lvert\frac{E_S}{N_0}\right\rvert^j\right]^{-1}$	$\left[2 \cdot \left\lvert\frac{\overline{E_S^{(Rx)}}}{N_0}\right\rvert^j\right]^{-1}$	$\left[2 \cdot \left[r^j\right]^2 \left\lvert\frac{\overline{E_S^{(Rx)}}}{N_0}\right\rvert^j\right]^{-1}$
a^j	$\frac{\sqrt{E_S}}{N_0}$	$\left\lvert\frac{E_S}{N_0}\right\rvert^j$	$r^j \cdot \left\lvert\frac{\overline{E_S^{(Rx)}}}{N_0}\right\rvert^j$	$\left[r^j\right]^2 \left\lvert\frac{\overline{E_S^{(Rx)}}}{N_0}\right\rvert^j$

Table 5.2: MRC Weighting Factors for the Individual Channel Models.

N independent copies $\mathbf{y}^0, \ldots, \mathbf{y}^{N-1}$ of a symbol corrupted by mean free additive noise and a constant fading to a single copy $\mathbf{y}_{MRC}$. To do so, the individual symbols $\mathbf{y}^0, \ldots, \mathbf{y}^{N-1}$ have to be multiplicatively weighted by factors $a^0, \ldots, a^{N-1}$ and to be added, resulting in the MRC symbol (see Figure 5.2):

$$\mathbf{y}_{MRC} = \sum_{j=0}^{N-1} a^j \cdot \mathbf{y}^j.$$

The weighting factors a^j are chosen in such a way that they are **proportional to the deterministic signal part** and **inversely proportional to the noise variance** in $\mathbf{y}^j$, $j = 0, \ldots, N-1$:

$$a^j = K \cdot \frac{E\{\mathbf{y}^j\}}{VAR\{\mathbf{y}^j\}},$$

with K being an arbitrary constant (however identical for all weightings). Hence, the correct weighting factor is proportional to the signal-**amplitude**-to-noise-power-density ratio.

Table 5.2 summerizes the resulting weighting factors (for $K = \frac{1}{2}$) for the presented AWGN and Rayleigh channel models in dependency of the corresponding channel parameters.

After the weighting, each of the N signal $a^j \cdot y^j$ have proportional expected value

$$\begin{aligned} E\left\{a^j \cdot \mathbf{y}^j\right\} &= a^j \cdot E\left\{\mathbf{y}^j\right\} \\ &= K \cdot \frac{\left[E\left\{\mathbf{y}^j\right\}\right]^2}{VAR\{\mathbf{y}^j\}} \end{aligned}$$

and variance

$$\begin{aligned} VAR\left\{a^j \cdot \mathbf{y}^j\right\} &= \left(a^j\right)^2 \cdot VAR\left\{\mathbf{y}^j\right\} \\ &= K^2 \cdot \frac{\left[E\left\{\mathbf{y}^j\right\}\right]^2}{VAR\{\mathbf{y}^j\}}. \end{aligned}$$

Hence, the MRC symbol has an expected value of

$$\begin{aligned} E\left\{\mathbf{y}_{MRC}\right\} &= \sum_{j=0}^{N-1} E\left\{a^j \cdot \mathbf{y}^j\right\} \\ &= K \cdot \sum_{j=0}^{N-1} \frac{\left[E\left\{\mathbf{y}^j\right\}\right]^2}{VAR\{\mathbf{y}^j\}} \end{aligned}$$

and a variance of

$$\begin{aligned} VAR\left\{\mathbf{y}_{MRC}\right\} &= \sum_{j=0}^{N-1} VAR\left\{a^j \cdot \mathbf{y}^j\right\} \\ &= K^2 \cdot \sum_{j=0}^{N-1} \frac{\left[E\left\{\mathbf{y}^j\right\}\right]^2}{VAR\{\mathbf{y}^j\}}. \end{aligned}$$

Consequently, the signal-energy-to-noise-power-density ratio of the MRC symbols is

$$\begin{aligned} \left.\frac{E_S}{N_0}\right|_{MRC} &= \frac{1}{2} \cdot \frac{[E\{\mathbf{y}_{MRC}\}]^2}{VAR\{\mathbf{y}_{MRC}\}} \\ &= \frac{K^2 \cdot \left[\sum_{j=0}^{N-1} \frac{\left[E\left\{\mathbf{y}^j\right\}\right]^2}{VAR\{\mathbf{y}^j\}}\right]}{2 \cdot K^2 \cdot \sum_{j=0}^{N-1} \frac{\left[E\left\{\mathbf{y}^j\right\}\right]^2}{VAR\{\mathbf{y}^j\}}} \\ &= \sum_{j=0}^{N-1} \frac{\left[E\left\{\mathbf{y}^j\right\}\right]^2}{2 \cdot VAR\{\mathbf{y}^j\}} \\ &= \sum_{j=0}^{N-1} \left.\frac{E_S}{N_0}\right|^j. \end{aligned} \tag{5.37}$$

Equation 5.37 proofs the statement, that MRC results in a symbol with the effective signal-energy-to-noise-power-density equal to the sum of all individual ratios (Equation 5.36). Without proof the result of Brennan is stated, that MRC is the method to obtain the highest effective $\frac{E_S}{N_0}$.

ML Decoding of a MRC Sequence vs. ML Decoding of the Individual Sequences

Brennan showed, that MRC yields the combined symbol with the highest signal-energy-to-noise-power-density ratio. Any other weighting results in a worse ratio. Hence, MRC is the best way of combining N copies of a symbol *prior* to decoding.

Channel Model	Equation	c_j	$\overrightarrow{a_j}$	$\overrightarrow{b}$
AWGN 1	5.25	$\frac{1}{N_0^j}$	$y^j[k]$	$x[k]$
AWGN 2	5.28	$\left\|\frac{E_S}{N_0}\right\|^j$	$y^j[k]$	$\overline{x}[k]$
Rayleigh 1	5.30	$\left\|\overline{\frac{E_S}{N_0}(Rx)}\right\|^j \cdot \left(r^j[k]\right)^2$	$\frac{y^j[k]}{r^j[k]}$	$\overline{x}[k]$
Rayleigh 2	5.32	$\left\|\overline{\frac{E_S}{N_0}(Rx)}\right\|^j \cdot \left(r^j[k]\right)^2$	$y^j[k]$	$\overline{x}[k]$
Multiple Rayleigh	5.34	$\left\|\overline{\frac{E_S}{N_0}(Rx)}\right\|^j[k]$	$y^j[k]$	$\overline{x}[k]$

Table 5.3: Corresponding Values in Equation 5.38 for the Individual Channel Models.

Now, if a decoder has available N mutual independent transmissions of a codeword we could apply

- ML decoding of multiple transmissions, as discussed in Section 5.1.2
- MRC to unify all received transmissions to a single one and use ML decoding of this single combined transmission (Section 5.1.1)

The questions which then arise are if both approaches yield the same results, i.e. if MRC followed by ML decoding of this combined transmission is another form of ML decoding of the individual transmissions and how is ML decoding of a maximum ratio combined transmissions performed.

In order to answer these questiones, we note that all equations, which describe the ML distance combining for repetitions (Equations 5.25 and 5.28 for the AWGN channel models, Equation 5.30 and 5.32 for the Rayleigh channel models, and Equation 5.34 for the Multiple Rayleigh channel model) are of the type

$$\sum_{j=0}^{N-1} c_j \cdot \left|\overrightarrow{a_j} - \overrightarrow{b}\right|^2 . \tag{5.38}$$

Herein c_j is a positive ($c_j > 0$) scalar and $\overrightarrow{a_j}$ as well as $\overrightarrow{b}$ are two dimensional vectors (complex numbers). Table 5.3 summerizes the corresponding values for the individual channel models.

Then, in Appendix C.7 it is shown that the following equation holds

$$\sum_{j=0}^{N-1} c_j \cdot \left|\overrightarrow{a_j} - \overrightarrow{b}\right|^2 = \frac{1}{\left(\sum_{j=0}^{N-1} c_j\right)} \cdot \left|\sum_{j=0}^{N-1} c_j \cdot \overrightarrow{a_j} - \left(\sum_{j=0}^{N-1} c_j\right) \cdot \overrightarrow{b}\right|^2 + f\left(\overrightarrow{a}_0, \ldots, \overrightarrow{a}_{N-1}, c_0, \ldots, c_{N-1}\right). \tag{5.39}$$

Hence, ML distance combining, (left side of Equation 5.39) can be written as a single squared distance computation, multiplied by a factor

$$\frac{1}{\left(\sum_{j=0}^{N-1} c_j\right)} > 0$$

Channel	$y_{MRC}[k]$	Squared Euclidean Distance
AWGN 1	$\sum_{j=0}^{N-1} \frac{1}{N_0^j} \cdot y^j[k]$	$\left\| y_{MRC}[k] - \left(\sum_{j=0}^{N-1} \frac{1}{N_0^j} \right) \cdot x[k] \right\|^2$
AWGN 2	$\sum_{j=0}^{N-1} \left\frac{E_S}{N_0} \right\|^j \cdot y^j[k]$	$\left\| y_{MRC}[k] - \left(\sum_{j=0}^{N-1} \left\frac{E_S}{N_0} \right\|^j \right) \cdot \overline{x}[k] \right\|^2$
Ray. 1	$\sum_{j=0}^{N-1} \left\overline{\frac{E_S}{N_0}(Rx)} \right\|^j \cdot r^j[k] \cdot y^j[k]$	$\left\| y_{MRC}[k] - \left(\sum_{j=0}^{N-1} \left\overline{\frac{E_S}{N_0}(Rx)} \right\|^j \cdot \left(r^j[k]\right)^2 \right) \cdot \overline{x}[k] \right\|^2$
Ray. 2	$\sum_{j=0}^{N-1} \left\overline{\frac{E_S}{N_0}(Rx)} \right\|^j \cdot \left(r^j[k]\right)^2 \cdot y^j[k]$	$\left\| y_{MRC}[k] - \left(\sum_{j=0}^{N-1} \left\overline{\frac{E_S}{N_0}(Rx)} \right\|^j \cdot \left(r^j[k]\right)^2 \right) \cdot \overline{x}[k] \right\|^2$
M. Ray.	$\sum_{j=0}^{N-1} \left\frac{E_S}{N_0} \right\|^j [k] \cdot y^j[k]$	$\left\| y_{MRC}[k] - \left(\sum_{j=0}^{N-1} \left\overline{\frac{E_S}{N_0}(Rx)} \right\|^j [k] \right) \cdot \overline{x}[k] \right\|^2$

Table 5.4: Squared Euclidean Distances to be Minimized for ML Decoding of MRC Sequences.

and an addend

$$f\left(\overrightarrow{a}_0, \ldots, \overrightarrow{a}_{N-1}, c_0, \ldots, c_{N-1}\right).$$

The factor, as well as the addend, are independent of the symbol $\overrightarrow{b}$ to which the distances are computed. Hence, a minimization of the left side of Equation 5.39 over $\overrightarrow{b}$ is identical to a minimization of the squared Euclidian distance

$$\left\| \sum_{j=0}^{N-1} c_j \cdot \overrightarrow{a_j} - \left(\sum_{j=0}^{N-1} c_j \right) \cdot \overrightarrow{b} \right\|^2 \tag{5.40}$$

over $\overrightarrow{b}$. Furthermore, the sum in Equation 5.40 is nothing but the correct MRC symbol (compare with Table 5.2) and Equation 5.40 states the squared distance which is required to be minimized for ML decoding of a MRC sequences. Table 5.4 summarizes the proper MRC equation and the corresponding ML distance computation for all presented channel models.

Hence, MRC with proper ML decoding yields the same results as ML distance combining. Since MRC combines all transmissions to a single one, only memory for one receive sequence per information sequence has to be provided. Therefore, for each symbol of the transmitted channel sequence 2 softdecision values (for the case of a two dimensional modulation alphabet) have to be stored, resulting in a maximum receiver memory requirement per information sequence of

$$M_p^{Rx}\big|_{max} = 2 \cdot L\,. \tag{5.41}$$

As an example, if a 16 PSK modulation alphabet is used, MRC combining requires 2 softdecision values per symbol, whereas distance combining requires 16. Clearly, if repetitions are sent, MRC is the preferable way of combining. However, the author likes to emphasis, that MRC is only possible *if* repetitions are sent and possible gains which might arise from changing the mapper (see next section) can not be obtained.

5.1.5 Examples

MRC

In Chapter 4, the rejection probabilities (Figure 4.5 on Page 58), the average number of transmissions (Figure 4.13 on Page 71), and the throughput (Figure 4.21 on Page 82) of the same CE-MARQ system were shown. This ARQ system makes use of a rate $\frac{1}{3}$ convolutional code with a constraint length $v = 8$ and uses BPSK as modulation scheme. For the first transmission, $\frac{2}{3}$ of the encoded bits are punctured, leading to a coding-/modulation rate of almost one[2]. The second transmission transmits the half of the previously punctured bits, yielding an overall coding-/modulation rate of $R_{CM}^{ARQ,2} = \frac{2}{3}$, and the third transmission delivers the remaining punctured bits. Hence, after 3 transmission the mother code rate is reached and the ARQ system has on overall coding-/modulation rate of $R_{CM}^{ARQ,3} = \frac{1}{3}$. For all further transmissions, the ARQ system transmits repetitions of the mother codeword and the decoder uses MRC for their combining. The effect of MRC is reflected in the rejection probability vs. $\frac{E_S}{N_0}$ plot (Figure 4.5). The separation between $P(R_3)$ and $P(R_4)$ is exactly $3dB$, representing the improvement of the SNR by the factor 2. In the same way, the .separation between $P(R_3)$ and $P(R_5)$ is $4.78dB$, resulting from a SNR improvement by the factor 3, and so on.

Distance Combining with Varying Mapper

In Section 5.1.3, ML distance combining was investigated. If the mapper remains identical for all transmissions, however, MRC yields the same performance with less memory requirement. Each new repetition leads to scaling of the distance profile, without changing its general shape.

For coding schemes which are followed by multilevel modulation, on the other hand, changing the mapper effectively results in a new code. Then, the code, which results from two combined transmission, has a new distance profile. If the sequence of mappings is chosen appropriately, the distance profile is improved. For TCM schemes, Schmitt [Schm98][Schm99], investigated techniques for deriving a set of two mapper. Figure 5.3 represents this procedure for the case of 8-PSK. It was shown, that this mapper set in conjunction with convolutional encoders, designed according to the Ungerboeck [Ung82] rules, yields to substantial performance gains over MRC at hardly any expense. Mayer [May98] used this technique in addition to puncturing on a symbol basis for further system improvements.

Figure 5.4 depicts his results for trellis coded 8-PSK with a constraint length $v = 8$ in comparison to the same system with constant mapper and MRC, and the channel capacity over an AWGN channel. It can be seen, that at $\frac{E_S}{N_0} = 6\,dB$ the system with varying mapper has a throughput improvement of 50 % over the system with constant mapper.

[2] The coding-/modulation rate is slightly smaller than 1, due to the flushing bits, which are required to bring the encoder in a known state at the end of the transmission. This deviation is neglected from now on.

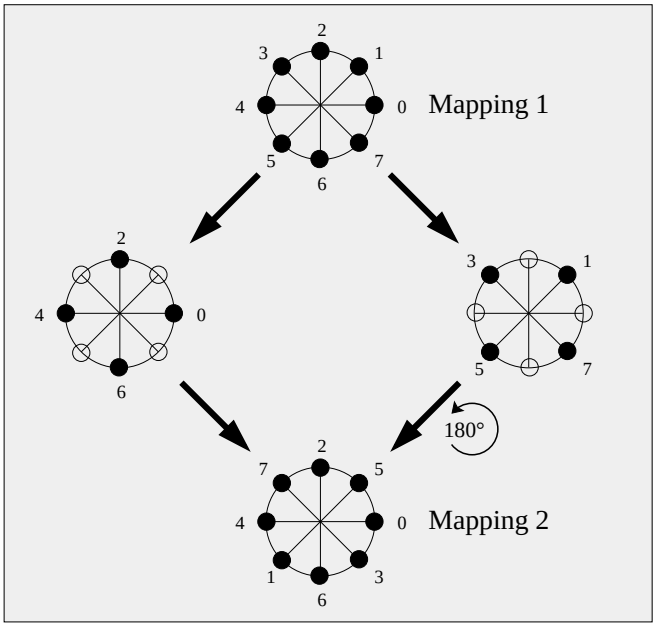

Figure 5.3: Rearrangement for Trellis Coded 8 PSK as shown in [Schm98].

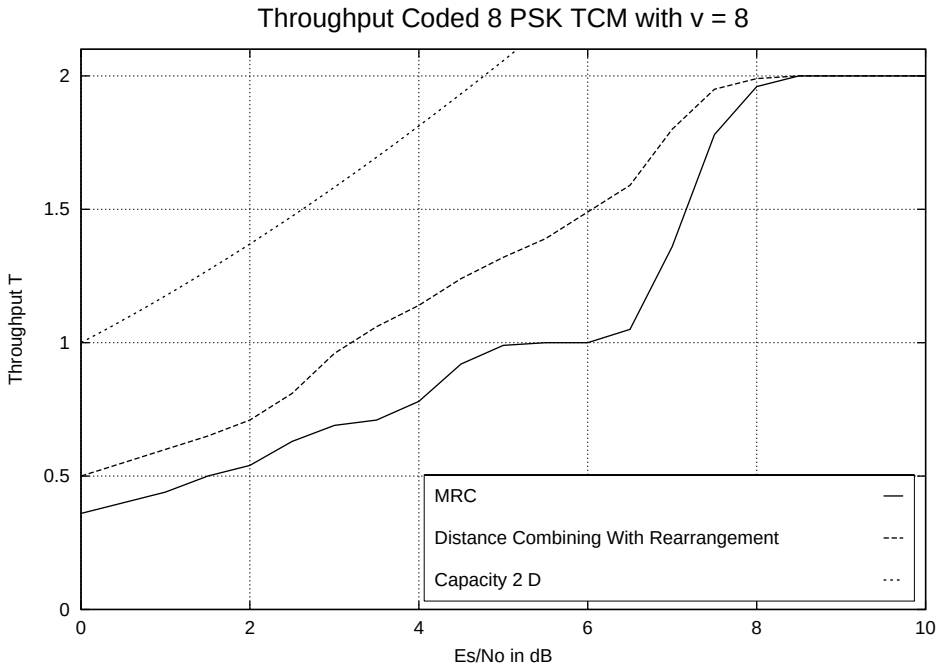

Figure 5.4: Throughput for a Coded 8 PSK TCM Scheme with MRC and with Distance Combining and Rearrangement.

5.2 Maximum A Posteriori Decoding

Throughout Section 5.1, the ML decoding of multiple transmission was discussed. In general for ML decoding, the valid channelword $\underline{x}$, which maximizes the conditioned probability density function[3]

$$f_{\underline{\mathbf{y}}|\underline{\mathbf{x}}}\left(\underline{y}\,|\underline{x}\right),$$

has to be found, i.e. the channelword $\underline{x}$, which makes the observed receiveword $\underline{y}$ most likely. In this section, the so-called maximum a posteriori (MAP) decoding is treated. Contrary to the ML decoding, the aim of MAP decoding is to determine the channelword $\underline{x}$, which is most likely if the receiveword $\underline{y}$ is observed, i.e. the probability[4]

$$P_{\underline{\mathbf{x}}|\underline{\mathbf{y}}}\left(\underline{x}\,|\underline{y}\right)$$

needs to be maximized. An algorithm for MAP decoding of a convolutional code is the so-called BCJR algorithm, named after its inventors Bahl, Cocke, Jelinek, and Raviv [BCJR74]. The input to the BJCR algorithm are the probabilities of the output bit sequences associated with the possible trellis transitions. If a rate $\frac{1}{n}$ code is used in conjunction with a binary modulation, n bit probabilities have to be calculated and properly combined to the various trellis transition probabilities. On the other hand, if a 2^n-ary modulation is used, the appropriate symbol probabilities are already the transition probabilities. However, if a 2^m-ary modulation is used together with a convolutional code of rate $\frac{1}{n}$ and $n \neq m$, the need arises to calculate the m individual bit probabilities in order to combine blocks of n bit probabilities to obtain the transition probabilities. Also, as it will be seen, the calculation of the bit error probabilities is essential if multilevel modulation is used to transmit turbo codes.

Again, let us consider the digital 2^m- ary modulation system depicted in Figure 5.1 on Page 123. Blocks b_c of m encoded bits

$$b_c = (b_0, b_1, \ldots, b_{k-1}) \in \{0,1\}^m$$

are mapped onto the channel symbol labels

$$c \in \{0, 1, \ldots, M-1\}$$

by bijective mapping, performed by the channel symbol label mapper. $M = 2^m$ denotes the number of input symbols and channel symbol labels.

The modulator maps the channel symbol labels c via the bijective mapping g_m onto the channel symbols x, which are represented by their complex signal space representation:

$$x = g_m(c) \in \{x_0, x_1, \ldots, x_{M-1}\},\ x_i \in C$$

[3] For continuous channel models the probability of observing a certain receiveword $\underline{y}$ is zero for all possible receivewords. Hence, the maximization must take place over the probability density function.

[4] Due to the finite size of the channelword alphabet, the probability of a certain channelword is non-vanishing and the maximization can be made over the probability.

Since the blocks of input bits b are mapped bijective onto the channel symbol labels c, which again are mapped onto the channel symbols s by a bijective mapping, it exists a bijective mapping g_b, which maps the sequence of encoded bits $\underline{b_c}$ onto the channel symbol sequence $\underline{x}$ on a symbols basis:

$$x = g_b\left(b_c\right). \tag{5.42}$$

Then, the channel symbol sequence $\underline{y}$ is passed through one of the discussed channel models, yielding the receive sequence $\underline{y}$. Again, we assume that the CSI is known perfectly. The receive sequence as well as the perfect CSI is used by the demodulator to compute the metrics, which are required by the decoder for the decoding process.

As said, for the BJCR algorithm, these metrics are the probabilities of transitions in the trellis under the constraint of the received symbol sequence.

If a transition consists of an integer number of transmission symbols, the probabilities of the received symbols $P_{\mathbf{y}|\mathbf{r}}\left(y|r\right)$ are required to compute these transition probabilities. This is always the case if binary modulation is used. In the case of M-ary modulation, with $M = 2^k$, these symbol probabilities are required, if a rate $\frac{1}{n \cdot k}$, $n \geq 1$ has to be decoded with the BCJR algorithm. Such examples are rate $\frac{1}{3}$codes and 8ary PSK modulation, or rate $\frac{1}{4}$ codes and QPSK modulation. The problem of computing these symbol probabilities for a single transmission is treated in Section 5.2.1.

On the other hand, if the transitions are not composed of integer numbers of channel symbols (for example in the case of a rate $\frac{1}{3}$ code in conjunction with a 16ary QAM), the symbol probabilities can not be used for the computation of the transition probabilities. In this case, the individual bit probabilities $P_{\mathbf{b}_i|\mathbf{y}}\left(b_i\,|y\right)$ have to be computed and with their help the transition probability is determined. The transmission of Turbo Codes [BGT93] with multilevel modulation represents another case, where this procedure must be applied. The reason for that can be found in the structure of the turbo codeword, which consists of encoder input bits, the coded bits of systematic rate $\frac{1}{2}$ convolutional code, and the encoded bits of a second systematic rate $\frac{1}{2}$convolutional code. No matter how these bits are arranged for transmission, as soon as multilevel modulation is applied, the individual bit probabilities are required for the decoding procedure. The problem of computing these bit probabilities for a single transmission is treated in Section 5.2.2.

Then, Section 5.2.3 investigates the simplified computation of these bit probabilities for regular QAM modulation schemes and, Section 5.2.4 moves on to the problem of how to determine these symbol and bit probabilities for multiple transmissions. As in the case of ML decoding, we allow for a varying mapper for the retransmissions.

Finally, Section 5.2.5 concludes this Chapter with the presentation of sample mapper rearrangements for MAP decoding and shows the possible throughput improvements.

5.2.1 Symbol Probabilities

Based on the observed continuous channel output y, the probability that a specific channel symbol label c was generated by the channel symbol label mapper shall be computed. That is, we want

to find $P_{\mathbf{c}|\mathbf{y}}(c|y)$ for all possible channel symbol labels c. Bayes Theorem for the above case states (see Appendix C.8.1, Page 198)

$$P_{\mathbf{c}|\mathbf{y}}(c|y) = \frac{f_{\mathbf{y}|\mathbf{c}}(y|c) \cdot P_{\mathbf{c}}(c)}{\sum_{j=0}^{M-1} f_{\mathbf{y}|\mathbf{c}}(y|j) \cdot P_{\mathbf{c}}(j)} \tag{5.43}$$

and with the assumption that the input bits b_i and hence all channel symbol labels c are equally likely,

$$P_{\mathbf{c}}(c) = \frac{1}{M} \; \forall c \in \{0, 1, \ldots, M-1\},$$

it follows

$$P_{\mathbf{c}|\mathbf{y}}(c|y) = \frac{f_{\mathbf{y}|\mathbf{c}}(y|c)}{\sum_{j=0}^{M-1} f_{\mathbf{y}|\mathbf{c}}(y|j)}. \tag{5.44}$$

The conditioned PDFs $f_{\mathbf{y}|\mathbf{c}}(y|c)$ have already been computed for the various channel models (Equation 3.6 for AWGN channel model 1, Equation 3.9 for AWGN channel model 2, Equation 3.14 for Rayleigh channel model 1, Equation 3.16 for Rayleigh channel model 2, and Equation 3.20 for the Multiple Rayleigh channel model). Their substitution into Equation 5.44 yields

$$P_{\mathbf{c}|\mathbf{y}}(c|y) = \frac{e^{-\frac{1}{N_0} \cdot |y - g_m(c)|^2}}{\sum_{j=0}^{M-1} e^{-\frac{1}{N_0} \cdot |y - g_m(j)|^2}} \tag{5.45}$$

for AWGN model 1,

$$P_{\mathbf{c}|\mathbf{y}}(c|y) = \frac{e^{-\frac{E_S}{N_0} \cdot |y - g_m(c)|^2}}{\sum_{j=0}^{M-1} e^{-\frac{E_S}{N_0} \cdot |y - g_m(j)|^2}} \tag{5.46}$$

for AWGN model 2,

$$P_{\mathbf{c}|\mathbf{y},\mathbf{r}}(c|y,r) = \frac{e^{-\overline{\frac{E_S^{(Rx)}}{N_0}} \cdot |y - r \cdot \overline{g_m}(c)|^2}}{\sum_{j=0}^{M-1} e^{-\overline{\frac{E_S^{(Rx)}}{N_0}} \cdot |y - r \cdot \overline{g_m}(j)|^2}} \tag{5.47}$$

for Rayleigh model 1,

$$P_{\mathbf{c}|\mathbf{y},\mathbf{r}}(c|y,r) = \frac{e^{-r^2 \cdot \overline{\frac{E_S^{(Rx)}}{N_0}} \cdot |y - \overline{g_m}(c)|^2}}{\sum_{j=0}^{M-1} e^{-r^2 \cdot \overline{\frac{E_S^{(Rx)}}{N_0}} \cdot |y - \overline{g_m}(j)|^2}} \tag{5.48}$$

for Rayleigh model 2, and

$$P_{\mathbf{c}|\mathbf{y},\frac{\mathbf{E_S}}{\mathbf{N_0}}}\left(c|y, \frac{E_S}{N_0}\right) = \frac{e^{-\frac{E_S}{N_0} \cdot |y - \overline{g_m}(c)|^2}}{\sum_{j=0}^{M-1} e^{-\frac{E_S}{N_0} \cdot |y - \overline{g_m}(j)|^2}}, \tag{5.49}$$

with $\frac{E_S}{N_0}$ representing the effective signal-energy-to-noise-power-density ratio, for the Multiple Rayleigh channel model.

5.2.2 Bit Probabilities

With the above derived Equations 5.45 to 5.49, the M probabilities, that a specific channel symbol c was generated by the mapper under the constraint that the receive symbol y was observed, can be calculated for the various channel models. A different equation need to be found if we are interested in the m probabilities $P_{\mathbf{b}_i|\mathbf{y}}\left(b_i\,|y\right)$, that the i-th bit in the transmitted bit block b has the value b_i, under the condition that the receive symbol y was observed.

Let $\underline{b^{(i)}}$ denote the set of all $b \in \{0,1\}^m$ with the i-th bit equal to one, i.e.

$$\underline{b^{(i)}} = \left\{b = (b_0, \ldots, b_{m-1}) \in \{0,1\}^m \,|b_i = 1\right\},\ i = 0, 1, \ldots, m-1,$$

and $\overline{\underline{b^{(i)}}}$ shall denote the set of all $b \in \{0,1\}^m$ with the i-th bit equal to zero, i.e.

$$\overline{\underline{b^{(i)}}} = \left\{b = (b_0, \ldots, b_{k-1}) \in \{0,1\}^m \,|b_i = 0\right\},\ i = 0, 1, \ldots, m-1.$$

Hence, $\underline{b^{(i)}}$ and $\overline{\underline{b^{(i)}}}$ provide a set partitioning of $\{0,1\}^m$, i.e.

$$\begin{aligned} \underline{b^{(i)}} \cup \overline{\underline{b^{(i)}}} &= \{0,1\}^m \\ \underline{b^{(i)}} \cap \overline{\underline{b^{(i)}}} &= \emptyset. \end{aligned}$$

Then, the probability that the i-th bit in b has a value one, under the condition that y was observed, is equal to the probability that b is element of $\underline{b^{(i)}}$, also under the condition that y was observed:

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = P_{\mathbf{b}|\mathbf{y}}\left(\underline{b^{(i)}}\,|y\right).$$

Since the events are mutually exclusive, the conditioned probability of the set is identical to the sum of over the conditioned probabilities of the set elements:

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \sum_{b \in \underline{b^{(i)}}} P_{\mathbf{b}|\mathbf{y}}\left(b\,|y\right).$$

Again, Bayes Theorem and the assumption of equal probabilities for the input bits yields

$$\begin{aligned} P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) &= \sum_{b \in \underline{b^{(i)}}} \frac{f_{\mathbf{y}|\mathbf{b}}\left(y|b\right)}{\sum_{\widetilde{b} \in \underline{b}} f_{\mathbf{y}|\mathbf{b}}\left(y\middle|\widetilde{b}\right)} \\ &= \frac{\sum_{b \in \underline{b^{(i)}}} f_{\mathbf{y}|\mathbf{b}}\left(y|b\right)}{\sum_{b \in \underline{b}} f_{\mathbf{y}|\mathbf{b}}\left(y|b\right)}. \end{aligned} \tag{5.50}$$

Due to the bijective mapping g_b (Equation 5.42), it follows for the conditioned PDF

$$f_{\mathbf{y}|\mathbf{b}}\left(y\,|b\right) = f_{\mathbf{y}|\mathbf{x}}\left(y\,|g_b\left(b\right)\right) \tag{5.51}$$

and substituting Equation 5.51 into 5.50 yields

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \frac{\sum_{b\in\underline{b}^{(i)}} f_{\mathbf{y}|\mathbf{x}}\left(y\,|g_b\left(b\right)\right)}{\sum_{b\in\underline{b}} f_{\mathbf{y}|\mathbf{x}}\left(y\,|g_b\left(b\right)\right)}. \tag{5.52}$$

Finally, substitutions of Equations 3.6, 3.9, 3.14, 3.9, and 3.20 in Equation 5.52 yield the conditioned bit probability for the various channel models. For the AWGN model 1

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \frac{\sum_{b\in\underline{b}^{(i)}} e^{-\frac{1}{N_o}\cdot|y-g_b(b)|^2}}{\sum_{b\in\underline{b}} e^{-\frac{1}{N_o}\cdot|y-g_b(b)|^2}} \tag{5.53}$$

is obtained, for the AWGN model 2

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \frac{\sum_{b\in\underline{b}^{(i)}} e^{-\frac{E_S}{N_o}\cdot|y-g_b(b)|^2}}{\sum_{b\in\underline{b}} e^{-\frac{E_S}{N_o}\cdot|y-g_b(b)|^2}} \tag{5.54}$$

is obtained, for the Rayleigh model 1

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y,r\right) = \frac{\sum_{b\in\underline{b}^{(i)}} e^{-\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot|y-r\cdot g_b(b)|^2}}{\sum_{b\in\underline{b}} e^{-\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot|y-r\cdot g_b(b)|^2}} \tag{5.55}$$

is obtained, for Rayleigh model 2

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y,r\right) = \frac{\sum_{b\in\underline{b}^{(i)}} e^{-r^2\cdot\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot|y-g_b(b)|^2}}{\sum_{b\in\underline{b}} e^{-r^2\cdot\frac{\overline{E_S^{(Rx)}}}{N_0}\cdot|y-g_b(b)|^2}} \tag{5.56}$$

is obtained, and finally, for the Multiple Rayleigh model we obtain

$$P_{\mathbf{b}_i\left|\mathbf{y},\frac{\mathbf{E_S}}{\mathbf{N_0}}\right.}\left(1\left|y,\frac{E_S}{N_0}\right.\right) = \frac{\sum_{b\in\underline{b}^{(i)}} e^{-\frac{E_S}{N_0}\cdot|y-g_b(b)|^2}}{\sum_{b\in\underline{b}} e^{-\frac{E_S}{N_0}\cdot|y-g_b(b)|^2}}. \tag{5.57}$$

The probability of the complementary event is given by

$$P_{\mathbf{b}_i|\mathbf{r}}\left(0\,|r\right) = \frac{\sum_{b\in\overline{\underline{b}^{(i)}}} e^{-\frac{1}{N_o}\cdot|y-g_b(b)|^2}}{\sum_{b\in\overline{\underline{b}}} e^{-\frac{1}{N_o}\cdot|y-g_b(b)|^2}} \tag{5.58}$$

or simply obtained via

$$P_{\mathbf{b}_i|\mathbf{y}}\left(0\,|y\right) = 1 - P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right).$$

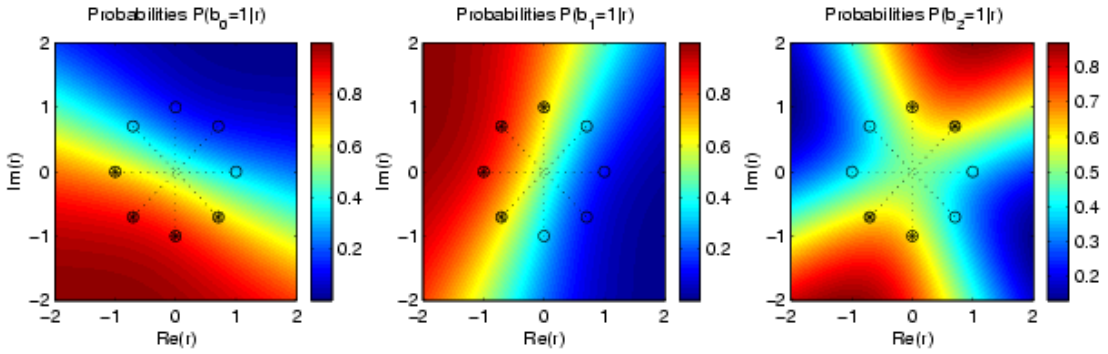

Figure 5.5: Conditioned Probabilities $P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right),\ i=0,\ldots,2$ for an 8-ary PSK Modulation

Finally, the the log-likelihood ratio for the bits

$$L_{\mathbf{b}_i|\mathbf{y}}\left(y\right)=ln\left(\frac{P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right)}{P_{\mathbf{b}_i|\mathbf{y}}\left(0\,|y\right)}\right)$$

is obtained for the various channel model via the substitution of Equations 5.53 to 5.57.

Figure 5.5 depicts the conditioned probabilities $P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right),\ i=0,\ldots,2$ for an 8-ary PSK modulation format and an effective $\frac{E_S}{N_0}=-2\ dB$. The filled circles indicate the symbols associated with the specific bits being one. It can be seen, that the conditioned PDFs are in general functions of 3 real variables: $\frac{E_S}{N_0}$, $Re\left\{y\right\}$, and $Im\left\{y\right\}$. In the following section, however, we consider specific modulation formats for which the number of dependent variables is reduced.

5.2.3 Simplified Computation of Bit Probabilities for Regular QAM

In general, the computation of the conditioned probabilities $P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right)$ and $P_{\mathbf{b}_i|\mathbf{r}}\left(0\,|r\right)$, as well as the log-likelihood ratio $L_{\mathbf{b}_i|\mathbf{y}}\left(y\right)$, requires $M=2^m$ exponential operations. However, for rectangular 2^n-ary QAM modulation with n even (i.e. 4, 16, 64, 256 QAM) and certain mappings g_b, a simplified computation requiring $2\cdot\sqrt{M}=2^{\frac{m}{2}+1}$ (i.e 4, 8, 16, 32) exponential operations is possible. Figure 5.6 depicts such a modulation format, the 16-ary QAM ($m=4$).

A rectangular 2^m-QAM, with m an even number, consists of $\sqrt{M}=2^{\frac{m}{2}}$ rows and columns, each with $2^{\frac{m}{2}}$ symbols. Let $\underline{s_R^{(i)}}$ and $\underline{s_C^{(j)}}$ denote the set of all symbols in the i-th row and the j-th column, respectively (see Figure 5.6). We are now interested in unions of $2^{\frac{m}{2}-1}$ different rows and unions of $2^{\frac{m}{2}-1}$ different columns. Therefore we define $\underline{p}$ and its complement to be an equal size partition of the set of all row and column numbers

$$\underline{p}\cup\overline{\underline{p}}=\left\{0,\ldots,2^{\frac{m}{2}}-1\right\}\ \wedge\ \left|\underline{p}\right|=\left|\overline{\underline{p}}\right|=2^{\frac{m}{2}-1}.$$

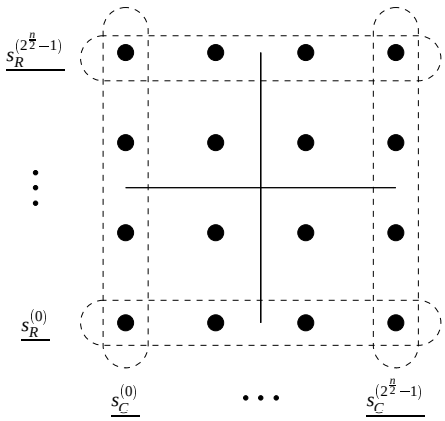

Figure 5.6: Rectangular 16-ary QAM.

For our example, $\underline{p} = \{0,1\}$ or $\underline{p} = \{1,3\}$ result in such partitions. Hence, the set given by the union of $2^{\frac{m}{2}-1}$ different rows

$$\underline{s_R}\left(\underline{p}\right) = \bigcup_{i \in \underline{p}} \underline{s_R^{(i)}}$$

is fully specified by $\underline{p}$. Of course, the same holds for the union of $2^{\frac{m}{2}-1}$ different columns $\underline{s_C}\left(\underline{p}\right)$.

It was already mentioned that the mapping g_b has to satisfy a certain property. As previously defined, let g_b be the bijective mapping of $\{0,1\}^m$ onto $\{0,\ldots,M-1\}$, this time, however, with the additional property that the m sets

$$\underline{b^{(i)}} = \{b = (b_0,\ldots,b_{m-1}) \in \{0,1\}^m \,|b_i = 1\}$$

are mapped onto either unions of rows $\underline{s_R}\left(\underline{p^{(i)}}\right)$ or unions of columns $\underline{s_C}\left(\underline{p^{(i)}}\right)$ with $\underline{p^{(i)}}$ and $\underline{\overline{p^{(i)}}}$ being equal size partitions of the row or column numbers. Figure 5.7 shows a 16-ary QAM modulation format with a mapping taken from a UMTS specification [UMTS98]. The bars labeled i_1, i_2, q_1, and q_2 indicate the regions where the various bits of the bit block $b = (i_1, q_1, i_2, q_2)$ are one. Clearly, the mapping satisfies the above mentioned constraints. With our notation, $\underline{p^{(0)}} = \underline{p^{(1)}} = \{0,1\}$, $\underline{p^{(2)}} = \underline{p^{(3)}} = \{0,3\}$, and bits b_0 and b_2 are mapped onto unions of columns, whereas bits b_1 and b_2 are mapped onto unions of rows.

Let x_j be the distance from the j-th column to the imaginary axes and, likewise, let y_j be the distance from the j-th row to the real axes. Then, as shown in Appendix C.8.2, for the mentioned assumptions the conditioned probability $P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right)$ is given by

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \begin{cases} \dfrac{\sum_{j \in \underline{p^{(i)}}} e^{-\frac{1}{N_o}\cdot(Re\{y\}-x_j)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-\frac{1}{N_o}\cdot(Re\{y\}-x_j)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_C} \\ \dfrac{\sum_{j \in \underline{p^{(i)}}} e^{-\frac{1}{N_o}\cdot(Im\{y\}-y_j)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-\frac{1}{N_o}\cdot(Im\{y\}-y_j)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_R} \end{cases} \tag{5.59}$$

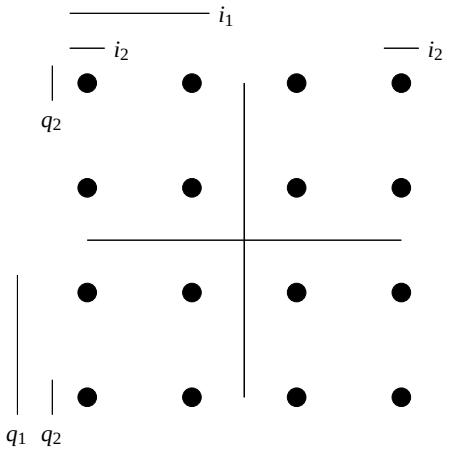

Figure 5.7: Rectangular 16-ary QAM with Mapping of the Bits $(b_0, b_1, b_2, b_3) = (i_1, q_1, i_2, q_2)$.

for the AWGN channel model 1, by

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \begin{cases} \dfrac{\sum_{j\in \underline{p^{(i)}}} e^{-\frac{E_S}{N_o}\cdot\left(Re\{y\}-x_j\right)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-\frac{E_S}{N_o}\cdot\left(Re\{y\}-x_j\right)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_C} \\ \\ \dfrac{\sum_{j\in \underline{p^{(i)}}} e^{-\frac{E_S}{N_o}\cdot\left(Im\{y\}-y_j\right)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-\frac{E_S}{N_o}\cdot\left(Im\{y\}-y_j\right)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_R} \end{cases} \tag{5.60}$$

for the AWGN channel model 2, by

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \begin{cases} \dfrac{\sum_{j\in \underline{p^{(i)}}} e^{-\overline{\frac{E_S}{N_o}(Rx)}\cdot\left(Re\{y\}-r\cdot x_j\right)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-\overline{\frac{E_S}{N_o}(Rx)}\cdot\left(Re\{y\}-r\cdot x_j\right)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_C} \\ \\ \dfrac{\sum_{j\in \underline{p^{(i)}}} e^{-\overline{\frac{E_S}{N_o}(Rx)}\cdot\left(Im\{y\}-r\cdot y_j\right)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-\overline{\frac{E_S}{N_o}(Rx)}\cdot\left(Im\{y\}-r\cdot y_j\right)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_R} \end{cases} \tag{5.61}$$

for the Rayleigh channel model 1, by

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \begin{cases} \dfrac{\sum_{j\in \underline{p^{(i)}}} e^{-r^2\cdot\overline{\frac{E_S}{N_o}(Rx)}\cdot\left(Re\{y\}-x_j\right)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-r^2\cdot\overline{\frac{E_S}{N_o}(Rx)}\cdot\left(Re\{y\}-x_j\right)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_C} \\ \\ \dfrac{\sum_{j\in \underline{p^{(i)}}} e^{-r^2\cdot\overline{\frac{E_S}{N_o}(Rx)}\cdot\left(Im\{y\}-y_j\right)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-r^2\cdot\overline{\frac{E_S}{N_o}(Rx)}\cdot\left(Im\{y\}-y_j\right)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_R} \end{cases} \tag{5.62}$$

M	# Exponential Operations Equations 5.53 to 5.57.	# Exponential Operations Equations 5.59 to 5.63	# Tables
4	4	4	1
16	16	8	2
64	64	16	3
256	256	32	4

Table 5.5: Summary of Complexity Reductions for Various Rectangular M-ary QAM Schemes.

for the Rayleigh channel model 2, and by

$$P_{\mathbf{b}_i|\mathbf{y}}\left(1\,|y\right) = \begin{cases} \dfrac{\sum_{j\in p^{(i)}} e^{-\frac{E_S}{N_o}\cdot(Re\{y\}-x_j)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-\frac{E_S}{N_o}\cdot(Re\{y\}-x_j)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_C} \\ \dfrac{\sum_{j\in p^{(i)}} e^{-\frac{E_S}{N_o}\cdot(Im\{y\}-y_j)^2}}{\sum_{j=0}^{2^{\frac{m}{2}-1}} e^{-\frac{E_S}{N_o}\cdot(Im\{y\}-y_j)^2}} & , if\ \underline{b^{(i)}} \mapsto \underline{s_R} \end{cases} \tag{5.63}$$

for the Multiple Rayleigh Channel model.

The implications of Equations 5.59 to 5.63 are twofold. Firstly, as already mentioned, they provide a reduction in the amount of exponential operations compared to Equations 5.53 to 5.57. These savings are summarized in Table 5.5. The second advantage concerns the arguments of the equations. Whereas Equations 5.53 to 5.57 have a complex arguments, namely the signal space representation of the received signal, their simplified counterparts only take a real number (either real or imaginary part of the received symbol). Figure 5.8 showing the conditioned probabilities $P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right),\ i=0,\ldots,3$ for the 16-ary QAM of Figure 5.7 and a $\frac{E_S}{N_0} = 2\,dB$.

The dependency on either the real or imaginary part only becomes immediately apparent (compare Figure 5.8 to 5.5). This fact could be used to replace the actual probability calculation by 2-dimensional look-up tables, where one dimension is $\frac{E_S}{N_0}$ and the second dimension becomes the real or imaginary part of the receive symbol. Hence, for an 2^n-ary QAM format, n even, in general n look-up tables are required. However, since the mapping of the bits associated with symbols in a row has no influence on the conditioned probability of the bits mapped onto symbols in columns, the same set of partitions can be used, leading to a reduction of the required number of look-up tables to $\frac{n}{2}$. The required number of two look-up tables is also shown in Table 5.5.

5.2.4 Multiple Transmissions

So far we considered one transmission only. We now want to investigate the case of multiple transmissions, i.e. the bit block b is transmitted N times with different effective signal-to-noise ratios $\frac{E_S}{N_0}$. In order to keep the retransmission scheme flexible, we allow different mappings g_c and hence different mappings g_b for each transmission. As in the previous sections, we are interested in the symbol probabilities and bit probabilities, however this time under the constraint that a series of receive symbols $y^j,\ j=0,\ldots,N$ were observed.

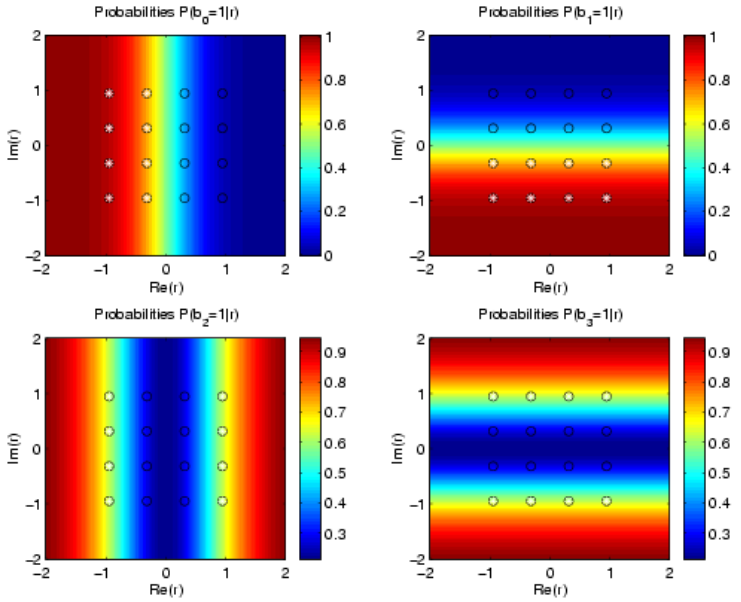

Figure 5.8: Conditioned Probabilities $P_{\mathbf{b}_i|\mathbf{r}}\left(1\left|r\right.\right),\ i=0,\ldots,3$ for the 16-ary QAM of Figure 5.8.

Again, let g_c^j denote the bijective mapping of c onto y for the j-th transmission. Correspondingly, g_b^j denotes the bijective mapping of b onto y.

The conditioned probabilities, we are interested in, are the probability of $\mathbf{c}=c$ under the constraint that $\mathbf{y}^0=y^0$, $\mathbf{y}^1=y^1$, ... , and $\mathbf{y}^{N-1}=y^{N-1}$ as well as the probability that the i-th bit in $\mathbf{b}=b$ has the value one, under the same constraints. These probabilities are denoted as $P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)$ and $P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)$, respectively.

For the **conditioned symbol probabilities**, application of Bayes Theorem yields

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{c}}\left(y^0,y^1,\ldots,y^{N-1}\left|c\right.\right)\cdot P_{\mathbf{c}}\left(c\right)}{\sum_{i=0}^{M-1}f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{c}}\left(y^0,y^1,\ldots,y^{N-1}\left|i\right.\right)\cdot P_{\mathbf{c}}\left(i\right)}$$

and with the assumption that all channel symbol labels are equally probable it follows

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{x}}\left(y^0,y^1,\ldots,y^{N-1}\left|c\right.\right)}{\sum_{i=0}^{M-1}f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{c}}\left(y^0,y^1,\ldots,y^{N-1}\left|i\right.\right)}. \tag{5.64}$$

Due to the mutual independence of the receive symbols it follows

$$f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{c}}\left(y^0,y^1,\ldots,y^{N-1}\left|c\right.\right)=\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{c}}\left(y^j\left|c\right.\right). \tag{5.65}$$

Substituting Equation 5.65 into Equation 5.64 yields

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{c}}\left(y^j\left|c\right.\right)}{\sum_{i=0}^{M-1}\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{c}}\left(y^j\left|i\right.\right)}. \tag{5.66}$$

Due to bijectivity of the mapper g_m^j we can rewrite Equation 5.66 as

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{x}}\left(y^j\left|g_m^j\left(c\right)\right.\right)}{\sum_{i=0}^{M-1}\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{x}}\left(y^j\left|g_m^j\left(i\right)\right.\right)}. \tag{5.67}$$

The exact equations for the various channel models can be obtain via a substitution of the appropriate Equation 3.6, 3.9, 3.14, 3.9, and 3.20 into Equation 5.67. Then, it follows

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{e^{-\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot\left|y^j-g_m^j(c)\right|^2}}{\sum_{i=0}^{M-1}e^{-\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot\left|y^j-g_m^j(i)\right|^2}} \tag{5.68}$$

for the AWGN model 1,

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{e^{-\sum_{j=0}^{N-1}\frac{E_S}{N_0}\left|^j\cdot\right|y^j-\overline{g}_m^j(c)\right|^2}}{\sum_{i=0}^{M-1}e^{-\sum_{j=0}^{N-1}\frac{E_S}{N_0}\left|^j\cdot\right|y^j-\overline{g}_m^j(i)\right|^2}} \tag{5.69}$$

for the AWGN channel model 2,

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{e^{-\sum_{j=0}^{N-1}\overline{\frac{E_S}{N_0}(Rx)}\left|^j\cdot\right|y^j-r^j\cdot\overline{g}_m^j(c)\right|^2}}{\sum_{i=0}^{M-1}e^{-\sum_{j=0}^{N-1}\overline{\frac{E_S}{N_0}(Rx)}\left|^j\cdot\right|y^j-r^j\cdot\overline{g}_m^j(i)\right|^2}} \tag{5.70}$$

for the Rayleigh channel model 1,

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{e^{-\sum_{j=0}^{N-1}(r^j)^2\cdot\overline{\frac{E_S}{N_0}(Rx)}\left|^j\cdot\right|y^j-\overline{g}_m^j(c)\right|^2}}{\sum_{i=0}^{M-1}e^{-\sum_{j=0}^{N-1}(r^j)^2\cdot\overline{\frac{E_S}{N_0}(Rx)}\left|^j\cdot\right|y^j-\overline{g}_m^j(i)\right|^2}} \tag{5.71}$$

for the Rayleigh channel model 2, and

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{e^{-\sum_{j=0}^{N-1}\frac{E_S}{N_0}\left|^j\cdot\right|y^j-\overline{g}_m^j(c)\right|^2}}{\sum_{i=0}^{M-1}e^{-\sum_{j=0}^{N-1}\overline{\frac{E_S}{N_0}}\left|^j\cdot\right|y^j-\overline{g}_m^j(i)\right|^2}} \tag{5.72}$$

for the Multiple Rayleigh channel model.

The important implication of Equations 5.68 to 5.72 is that for the symbol probability calculation after several transmissions of a channel symbol label c with possibly varying mappings $g_c^{(j)}$, simply the ML combined distances have to be used. As a consequence, if a constant mapping is

used, MRC could also be used as combining method prior to the probability calculation.

The author likes to emphasis that

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right) \neq \prod_{j=0}^{N-1} f_{\mathbf{c}|\mathbf{y}}\left(c\left|y^j\right.\right). \tag{5.73}$$

The right part of Equation 5.73, i.e. the product of the individual conditioned probabilities, presents the probability that the symbol c is transmitted multiple times with all combinations of symbols from $\underline{c}$ being possible. Hence,

$$P_{\mathbf{c}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(c\left|y^0,y^1,\ldots,y^{N-1}\right.\right) > \prod_{j=0}^{N-1} f_{\mathbf{c}|\mathbf{y}}\left(c\left|y^j\right.\right).$$

We now turn to the **bit probability** calculation for multiple transmissions of a bit block b. As previously, $\underline{b^{(i)}}$ denotes the subset of all b with the i-th bit equal to one. The probability that the i-th bit in b is one under the condition that the series of channel outputs y^j, $j = 0,\ldots,N-1$ are observed is equal to the probability that $b \in \underline{b^{(i)}}$ under the same constraints

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right) = P_{\mathbf{b}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(\underline{b^{(i)}}\left|y^0,y^1,\ldots,y^{N-1}\right.\right)$$

and also equal to the sum over the individual probabilities $b \in \underline{b^{(i)}}$ under the same constraints:

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right) = \sum_{b\in\underline{b^{(i)}}} P_{\mathbf{b}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(b\left|y^0,y^1,\ldots,y^{N-1}\right.\right). \tag{5.74}$$

From Bayes Theorem in conjunction with equal probability for the bits it follows

$$P_{\mathbf{b}|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(b\left|y^0,y^1,\ldots,y^{N-1}\right.\right) = \frac{f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{b}}\left(y^0,y^1,\ldots,y^{N-1}\left|b\right.\right)}{\sum_{\widetilde{b}\in\underline{b}} f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{b}}\left(y^0,y^1,\ldots,y^{N-1}\left|\widetilde{b}\right.\right)}. \tag{5.75}$$

Substituting Equation 5.75 into 5.74 leads to

$$\begin{aligned} P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right) &= \sum_{b\in\underline{b^{(i)}}} \frac{f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{b}}\left(y^0,y^1,\ldots,y^{N-1}\left|b\right.\right)}{\sum_{\widetilde{b}\in\underline{b}} f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{b}}\left(y^0,y^1,\ldots,y^{N-1}\left|\widetilde{b}\right.\right)} \\ &= \frac{\sum_{b\in\underline{b^{(i)}}} f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{b}}\left(y^0,y^1,\ldots,y^{N-1}\left|b\right.\right)}{\sum_{b\in\underline{b}} f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{b}}\left(y^0,y^1,\ldots,y^{N-1}\left|b\right.\right)} \end{aligned} \tag{5.76}$$

From the mutual independence of the $\mathbf{y}^j$, $j = 0,\ldots,N-1$ it follows

$$f_{\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}|\mathbf{b}}\left(y^0,y^1,\ldots,y^{N-1}\left|b\right.\right) = \prod_{j=0}^{N-1} f_{\mathbf{y}|\mathbf{b}}\left(y^j\left|b\right.\right) \tag{5.77}$$

and substituting Equation 5.77 into Equation 5.76 yields

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\sum_{b\in\underline{b}^{(i)}}\left[\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{b}}\left(y^j\left|b\right.\right)\right]}{\sum_{b\in\underline{b}}\left[\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{b}}\left(y^j\left|b\right.\right)\right]}. \tag{5.78}$$

Due to the bijective mapper g_b^j, we can rewrite Equation 5.78 as

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\sum_{b\in\underline{b}^{(i)}}\left[\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{x}}\left(y^j\left|g_b^j\left(b\right)\right.\right)\right]}{\sum_{b\in\underline{b}}\left[\prod_{j=0}^{N-1}f_{\mathbf{y}|\mathbf{x}}\left(y^j\left|g_b^j\left(b\right)\right.\right)\right]} \tag{5.79}$$

Again we can use the Equations 3.6, 3.9, 3.14, 3.9, and 3.20 in conjunction with Equation 5.79 to derive the exact equations for all introduced channel models. Then, we obtain

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\sum_{b\in\underline{b}^{(i)}}\left[e^{-\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot\left|y^j-g_b^j(b)\right|^2}\right]}{\sum_{b\in\underline{b}}\left[e^{-\sum_{j=0}^{N-1}\frac{1}{N_0^j}\cdot\left|y^j-g_b^j(b)\right|^2}\right]} \tag{5.80}$$

for the AWGN channel model 1,

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\sum_{b\in\underline{b}^{(i)}}\left[e^{-\sum_{j=0}^{N-1}\frac{E_S}{N_0}\left|^j\cdot\right|y^j-g_b^j(b)\right|^2}\right]}{\sum_{b\in\underline{b}}\left[e^{-\sum_{j=0}^{N-1}\frac{E_S}{N_0}\left|^j\cdot\right|y^j-g_b^j(b)\right|^2}\right]} \tag{5.81}$$

for the AWGN channel model 2,

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\sum_{b\in\underline{b}^{(i)}}\left[e^{-\sum_{j=0}^{N-1}\overline{\frac{E_S}{N_0}(Rx)}\left|^j\cdot\right|y^j-r^j\cdot g_b^j(b)\right|^2}\right]}{\sum_{b\in\underline{b}}\left[e^{-\sum_{j=0}^{N-1}\overline{\frac{E_S}{N_0}(Rx)}\left|^j\cdot\right|y^j-r^j\cdot g_b^j(b)\right|^2}\right]} \tag{5.82}$$

for the Rayleigh channel model 1,

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\sum_{b\in\underline{b}^{(i)}}\left[e^{-\sum_{j=0}^{N-1}(r^j)^2\cdot\overline{\frac{E_S}{N_0}(Rx)}\left|^j\cdot\right|y^j-g_b^j(b)\right|^2}\right]}{\sum_{b\in\underline{b}}\left[e^{-\sum_{j=0}^{N-1}(r^j)^2\cdot\overline{\frac{E_S}{N_0}(Rx)}\left|^j\cdot\right|y^j-g_b^j(b)\right|^2}\right]} \tag{5.83}$$

for the Rayleigh channel model 2, and finally

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)=\frac{\sum_{b\in\underline{b}^{(i)}}\left[e^{-\sum_{j=0}^{N-1}\frac{E_S}{N_0}\left|^j\cdot\left|y^j-g_b^j(b)\right|^2}\right]}{\sum_{b\in\underline{b}}\left[e^{-\sum_{j=0}^{N-1}\frac{E_S}{N_0}\left|^j\cdot\left|y^j-g_b^j(b)\right|^2}\right]} \tag{5.84}$$

for the Multiple Rayleigh channel model.

As for the symbol probabilities, Equations 5.80 to 5.84 imply that for the computation of the conditioned bit probabilities, the receive symbols have to be ML combined prior to the probability calculation. Again, the author like to emphasis that

$$P_{\mathbf{b}_i|\mathbf{y}^0,\mathbf{y}^1,\ldots,\mathbf{y}^{N-1}}\left(1\left|y^0,y^1,\ldots,y^{N-1}\right.\right)\neq\prod_{j=0}^{N-1}f_{\mathbf{b_i}|\mathbf{y}}\left(1\left|y^j\right.\right).$$

5.2.5 Rearrangements

In last section, the problem of calculating the individual bit probabilities under the condition that several transmissions of a channel symbol are available was treated. Throughout this analysis, we allowed for different mappers g_b^j. Like in the ML decoding, this freedom can be used to improve the performance of an ARQ system, which employs MAP decoding, such as a turbo coded ARQ system.

In the ML decoding case, the mapper rearrangements presented by Schmitt [Schm98] was derived from the inherent code structure in a TCM scheme, which has been designed according to the Ungerboeck [Ung82] rules. Due to this code structure, a rearrangement scheme could be presented, which yield considerable performance gains independent of the actual TCM scheme.

For turbo codes, on the other hand, the decoding is much more complex, involving iterative decoding procedures. The relation of the mapping of the codebits onto the channel symbols and the decoding performance cannot be analyzed, as it can be made in the case of Viterbi decoding. Hence, no code dependent rearrangement can be given.

Nevertheless, a closer look on Figure 5.8 reveals that due to the multilevel modulation the different bits (b_0,b_1,b_2,b_3) have on average a different MAP probability. To illustrate this, compare the plot of bit b_0 (upper left corner) and the plot of bit b_2 (lower left corner). Now assume that both bits are one. Then, for bit $b_0=1$, one of the symbol of the two most left columns is used for transmission (which one depends on the other three bits). During the transmission process, noise is added and a receive symbol with a certain deviation from the original transmitted symbol is received. If a symbol from the left-most column is chosen and if, by chance, no noise has been added, the MAP probability calculation yields a probability of roughly 92%. If a symbol from the second column has been chosen and if, again by chance, no noise has been added, a MAP probability of about 70% is obtained. Now, consider the bit $b_2=1$. Depending on the other bits, a symbol from two outer columns is chosen. If, by chance, no noise is added during the transmission process, the MAP probability for that bit is obtained to be roughly 62%. Although the case

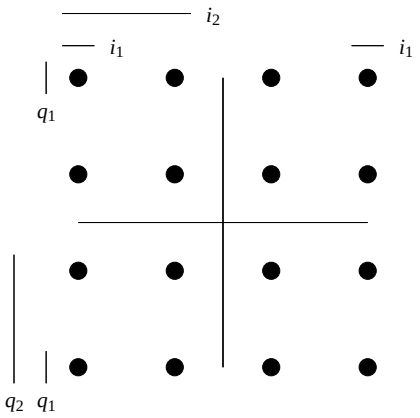

Figure 5.9: Rearranged 16-ary QAM Mapping of the Bits $(b_0, b_1, b_2, b_3) = (i_1, q_1, i_2, q_2)$ ..

of no noise is only a sample noise instant (with zero probability), it becomes clear that the bit b_0 has on average (that is, averaged over all possible noise samples) a higher MAP probability than bit b_2. Hence, due to the mapping, the bit b_2 is more sensitive to noise. The same argumentation holds (independently of b_0 and b_2) for the bits b_1 and b_3. As a conclusion, the bits b_2 and b_3 are the most likely sources of a decoding failure.

If a retransmission is made with the same mapping and MRC, the MAP probabilities for all bits are improved (leading to the 3 dB gain for AWGN channel), but the fact that the two bits b_2 and b_3 have on average worse MAP probabilities remains unchanged. If the second transmission also fails, it is most likely again due to this two more unreliable bits. Therefore, for an improved retransmission strategy, the MAP probability improvement with the second transmission for the two less reliable bits should be increased compared to MRC. This can be achieved via a mapper rearrangement. Therefore, compare a rearranged mapping of 16-ary QAM, depicted in Figure 5.9, with the original mapping of Figure 5.7. A comparison reveals, that in the rearrangement simply the mappings of the bits b_0 and b_2 and the mappings of the bits b_1 and b_3 are exchanged. Hence, the previously unreliable bits b_2 and b_3 are in the second transmission more reliable. After the combining, all four bits have on average the same average MAP probability, which is increased compared to the average MAP probability of the MRC scheme. Figure 5.10 depicts the throughput of the MRC combined system and the throughput of the distancecombined system, which alternatively employs the two mappings of Figure 5.7 and 5.9). It can be seen, that the simple procedure of mapper rearrangement yields 1 dB performance gain as soon as retransmissions become likely.

In the presented example of 16-ary QAM, an equal distribution of the average MAP probability was already achieved via a simple exchange of each two bits. In general, this equal distribution of the average MAP probability for all bits is not possible. Consider, for example, the 8-ary PSK modulation in Figure 5.5. From the three bits, b_2 has the lowest average MAP probability, whereas the other two bits have an identical average MAP probability. Clearly, for the second transmission, the mapping of bit b_2 has to be exchanged with either b_0 or b_1 to yield an performance

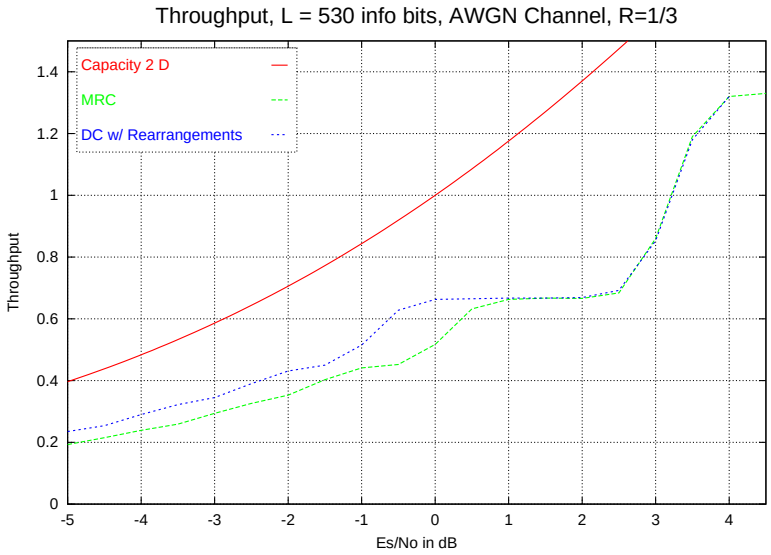

Figure 5.10: AWGN Channel Capacity and Throughput of Two Turbo Coded (Rate 1/3) Multilevel Modulation (16-ary QAM) ARQ Schemes.

improvement compared to MRC. Without loss of generality, we assume that b_1 is chosen and after the second transmission, the average MAP probability of b_1 and b_2 are identical and lower than the average MAP probability of b_0. Independent of what two bit mapping are exchanged next, no equal distribution of the average MAP probabilities can be achieved. However, their relative difference can be made smaller and smaller with a series of rearrangements.

A further observation is, that the relative difference between the average MAP probabilities for the individual bits, increases with the size of the modulation alphabet. Figure 5.11 depicts the MAP probabilities for a 64-ary QAM modulation format as specified in the UMTS specifications It can be seen that the bits b_0 and b_1 have the highest average MAP probability, whereas the bits b_4 and b_5 have the lowest. In the special case, where the channel adds no noise, these two bits have an average MAP probability of less than 60%. Accordingly, we expect the possible throughput gain for an rearrangement to be larger for a rearrangement of this modulation format, than for the 16-ary QAM. Figure 5.12 indeed shows a performance gain of more than 1.5 dB.

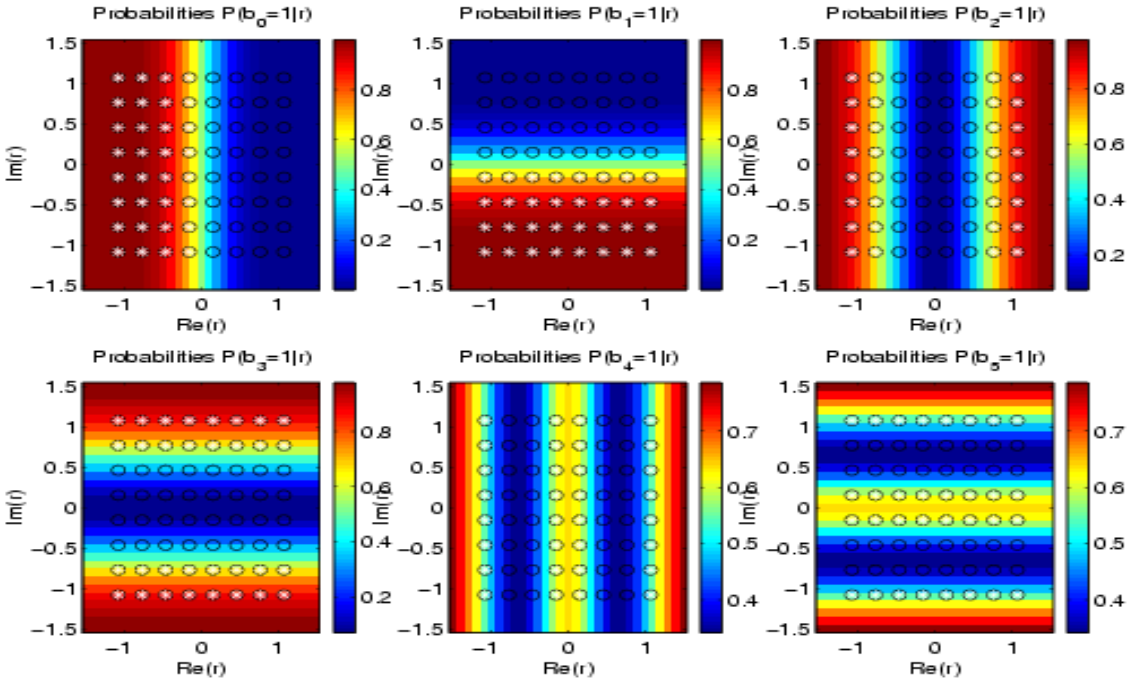

Figure 5.11: Conditioned Probabilities $P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right),\, i=0,\ldots,5$ for a 64-ary QAM.

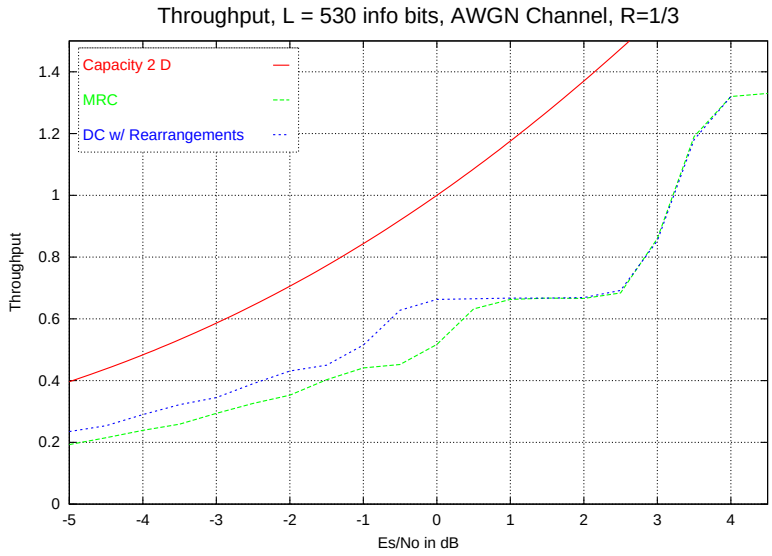

Figure 5.12: AWGN Channel Capacity and Throughput of Two Turbo Coded (Rate 1/3) Multilevel Modulation (64-ary QAM) ARQ Schemes.

Chapter 6

Summary and Outlook

In the following, a brief summary of this works is given and the novel contributions of this work are highlighted. Thereafter, a short outlook for related future work is given.

Summary

Throughout this work, a special class of feedback systems for digital transmission, namely ARQ systems, were investigated . Therefore, Chapter 1 provided an introduction by distinguishing these systems from forward error correction systems and showing their relation to the ISO-OSO model. It was argued, that a combined implementation of Layer 1 and 2 provides the basis for possible performance gains.

Chapter 2 aimed for a common ARQ system terminology. Therefore, Section 2.1 presented the classification of ARQ systems based on transmission protocol (SW, GBN, and SR protocol). These classification is well known and commonly agreed on, whereas Section 2.2 revealed, that further classifications of ARQ systems, which can be found in literature, are used inconsistently, and even worse, are based on specific decoder implementations. Therefore, Section 2.3 distinguished ARQ systems from general feedback systems and introduced a new concept of ARQ classification, which is solely based on the encoder protocol, i.e. what is sent. The way the decoder make use of the information is not part of this classification. The most basic destinction between decoding procedures is, if the decoder utilizes more than the currently received transmission for the decoding process, i.e. if the decoder employs memory. Therefore, Section 2.4 seized an old definition, which did not prevail in literature, for the classification of ARQ system implementations.

At the end of Chapter 2, a common terminology was established and the work continued with the presentation of various wireless channel models in Chapter 3. The reason for that in depth treatments can be found in the lack of literature dedicated to that subject and the problems which arise when several transmissions are to be ML or MAP combined. Also, the channel model statistics, which were derived for all introduced models, forming the basis for the investigations of optimum combining methods in Chapter 5.

Chapter 4, in which ARQ measures were introduced and analyzed, forms a major part of

this work. As a foundation of this treatment, Section 4.1 distinguishes the ratios $\frac{E_S}{N_0}$ and $\frac{E_b}{N_0}$ and extends the definition of the coding-/modulation rate of FEC systems to ARQ systems. Then, Section 4.2 bounds the retransmission probabilities of an ARQ system with the help of the rejection probabilities of an equivalent diversity scheme. These probabilities were used throughout the remainder of Chapter 4 to bound all newly defined performance measures. The first introduced performance measure was the number of transmission (Section 4.3). Contrary to all treatments in literature, found by the author, not only the average value, but also the general distribution was analyzed. In addition, as another contribution of this work, the treatment of this measure was extended to systems with non-constant retransmission probabilities. Section 4.4 introduced the throughput as the next performance measure. A new and meaningful definition was given, for which it was shown that the Shannon channel capacity provides an upper bound and that a functional relation between the throughput and $\frac{E_b}{N_0}$ not necessarily exists. Thereafter, the throughput performance was analyzed. Among the novel contributions of this work are the analysis of systems with memory and the throughput analysis for systems with variable packet sizes. Section 4.5 investigated an ARQ performance measure, which is hardly treated at all in literature, namely the delay. Therefore, it was differentiated between the packet and the information delay. The average values of these delays were analyzed for all three transmission protocols of Section 2.1. Then, Section 4.6 very generally investigated the transmitter and the receiver requirements. The performance measure chapter is concluded with a treatment of the data rate of an ARQ systems (Section 4.7). After the introduction of the theoretical limit it was shown that the data rate can be plotted in two ways. Then, the data rate is analyzed for all introduced transmission protocols.

Finally, the Chapter 5 treats the problem of optimal combining of retransmissions. Section 5.1 investigates the problem of ML combining of transmissions of the same codeword with possibly varying mappers. The results of Schmitt [Schm98] for ML decoding in the case of AWGN channel are extended to ML decoding of all introduced channel models. From this results, the ML distance combining is derived. As another contribution, it was shown that, in the case of a constant mapper, MRC is another form of ML combining and the appropriate ML weighting and distance computation was derived. Section 5.2 investigates the problem of MAP symbol and bit probability calculation. The results are obtained for all presented channel models. As a contribution, Section 5.2.3 reveals that the bit MAP probability calculation for regular QAM modulation can be simplified. Thereafter, a problem which has not been addressed in literature so far has been analyzed: The MAP combining of multiple transmission which consists of the same codeword, transmitted via a possibly varying mapper. Finally, the principle is shown, how to generate a second mapping so that an ARQ system, which uses this two mapping consecutively for transmission outperforms the corresponding MRC system. With this principle, a performance gain of about 1 to 1.5 dB can be obtained for the regions, where on average two transmissions are likely.

Outlook

With this work, the foundation for the analysis and the information-theoretical judgment of existing ARQ systems have been laid. Also the optimal ML and MAP combining methods were

derived. However, besides the the multilevel rearrangement scheme, which has been presented in Section 5.2.5, no novel ARQ schemes were presented. As a practical extension of this work, its results, or at least the insight which are obtained by applying this work to existing ARQ proposals, can be used to design novel ARQ systems which perform well - not only in information-theoretical terms, but also in practical comparisons.

As discussed in Chapter 2, ARQ systems are special part of feedback systems. Further work could be done to extend the performance analytical part to general information feedback systems, i.e. to provide analysis also for information feedback systems. A theoretically very interesting question, in this context, is the performance improvement of information feedback systems over ARQ systems if the error location is known perfectly. Also such a perfect error localization does not exist, it will in general show how much can be gained. For decision feedback systems also the definition of throughput must be extended, since the amount of information which is transmitted over the feedback channel is not negligible. Also, due to the increased feedback traffic, the assumption of an error free feedback channel can not be made and must be considered in the analysis.

Appendix A

List of Acronyms

ACF	Autocorrelation Function
ACK	Positive Acknowledgment
AGC	Automatic Gain Control
APP	A Posteriori Probability
ARQ	Automatic-Repeat-Request
ASK	Amplitude Shift Keying
AWGN	Additive White Gaussian Noise
BCJR	Bahl, Cocke, Jelinek, and Raviv [BCJR74]
BER	Bit Error Rate
BLER	Block Error Rate
BPSK	Binary Phase Shift Keying
CRC	Cyclic Redundancy Check
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CE-ARQ	Constant Encoder ARQ
CSI	Channel State Information
CSMA	Carrier Sense Multiple Access
DVB	Digital Video Broadcast
ETSI	European Telecommunications Standards Institute

FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
FM	Frequency Modulation
GBN	Go-Back-N
GMSK	Gaussian Minimum Shift Keying
HARQ	Hybrid Automatic-Repeat-Request
HEC	Header Error Check
HSDPA	High Speed Downlink Packet Access
IIR	Infinite Impulse Response
IP	Internet Protocol
ISI	Intersymbol Interference
ISO	International Standards Organization
LLC	Logical Link Control
LoS	Line-of-Sight
LTI	Linear Time Invariant
MAC	Media Access Control
MARQ	Memory ARQ
MAP	Maximum a Posteriori
ML	Maximum-Likelihood
MRC	Maximum-Ratio-Combining
NAK	Negative Acknowledgment
OFDM	Orthogonal Frequency Division Multiplexing
OSI	Open Systems Interconnection
PDF	Probability Density Function
PDU	Packet Data Unit
PSD	Power Spectrum Density
PSK	Phase Shift Keying

QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
RF	Radio Frequency
r.v.	Random Variable
R_x	Receive
SDU	Service Data Unit
SNR	Signal - to - Noise Ratio
SR	Selective-Repeat
SW	Stop-and-Wait
TCM	Trellis Coded Modulation
TCP	Transport Control Protocol
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
T_x	Transmit
UMTS	Universal Mobile Telecommunications System
VE-ARQ	Variable Encoder ARQ
WSSUS	Wide Sense Stationary Uncorrelated Scattering

Appendix B

List of Symbols

$\underline{b}$	Information Bit Sequence
b	Information Bit
$\underline{c}$	Encoded Bit Sequence
c	Encoded Bit
$\Delta t_I^{(N_I)}$	Information Frame Delay of a Frame of Size N_I
Δt_p	Packet Delay
$E\{\cdot\}$	Expected Value
$E_S^{(R_x)}$	Receive Signal Energy
$E_S^{(T_x)}$	Transmit Signal Energy
$\frac{E_b}{N_0}$	Normalized Energy to Noise Power Density Ratio
$\frac{E_S}{N_0}$	Signal Energy to Noise Power Density Ratio
f	Frequency
$f_{\cdot}(\cdot)$	Density Function
$f_{\cdot\|\cdot}(\cdot\|\cdot)$	Conditioned Density Function
$F_{\cdot}(\cdot)$	Cumulative Density Function
K^{GBN}	GBN Protocol Degradation Factor
K^{SR}	SR Protocol Degradation Factor
K^{SW}	SW Protocol Degradation Factor
l	Multipath Channel Diversity

L	Total Length of Packet in $symbol$
L_{info}	Length of Information to be Transmitted in bit
L_I	Information Frame Length
L_j	Length of Packet at the j-th Transmission
M	Size of the Modulation Alphabet
$M_p^{T_x}$	Packet Memory Requirement at the Transmitter
$M_{tot}^{T_x}$	Total Max Transmitter memory Requirement
N_0	Double Sided Noise Power Density
N_I	Number of Packets per Frame
N_{RT}	Round Trip Number
$\underline{\mathbf{n}}$	Sampled Noise Sequence
$\mathbf{n}$	Noise Sample
n_{RT}^j	Ratio of Round Trip Delay and j-th Channel Packet Duration
n_{trans}	Number of Transmissions
p_j	Multipath Tap Ratio
$\overline{p_j}$	Normalized Multipath Tap Ratio
$P(R_j)$	Rejection Probability of j- Diversity System
$P(RR_j)$	Retransmission Request Probability of the j-th Transmission n an ARQ System
P_S	Signal Power
$\frac{P_S}{N_0}$	Signal-Power-to-Noise-Power-Density Ratio
$\underline{\mathbf{r}}$	Fading Sequence
$\mathbf{r}$	Fading Instant
R_j	Event Rejection of the Combined Transmissions of a j-Diversity System
RR_j	Event Decoding of j-th Transmission and All Previous $j-1$ Transmissions Fails (in an ARQ system)
R_C	FEC Code Rate
$R_{CM}^{ARQ,j}$	Total Code-/Modulation Rate of an ARQ System after the j-th Transmission
R_{CM}^{FEC}	Code-/Modulation Rate of a FEC System

$R_{CM}^{FEC,j}$	Code-/Modulation Rate of the FEC System incorporated into an ARQ System at the j-th Transmission
SNR	Signal to Noise Ratio
σ	Standard Deviation
σ^2	Variance
T	Throughput
t_{prc}	Processing Time
t_{prp}	Propagation Time
t_{RT}	Round-Trip Delay
t_S	Channel Symbol Duration
v	Overall Constraint Length
$\underline{x}$	Channel Symbol Sequence
x	Channel Symbol
x_j	Possible Channel Symbol
$\underline{\mathbf{y}}$	Sampled Receive Sequence
$\mathbf{y}$	Receive Sample
$\lceil \cdot \rceil$	Rounding to Next Higher Integer Value

Appendix C

Mathematical Appendix

C.1 Rayleigh Channel Derivations

C.1.1 Generation of a χ^2 Distributed Random Variable

A χ^2 distributed random variable $\mathbf{x}$ of degree of freedom N can be generated by adding the square of N mutually independent $N(0,1)$ distributed random variables $\mathbf{n}_i$

$$\mathbf{x} = \sum_{i=1}^{N} \mathbf{n}_i^2.$$

The corresponding PDF is

$$f_{\mathbf{x}}(x) = \frac{x^{\frac{N}{2}-1}}{2^{\frac{N}{2}} \cdot \Gamma\left(\frac{N}{2}\right)} \cdot e^{-\frac{x}{2}} \cdot u(x)$$

and expected value are[1]

$$E\{\mathbf{x}\} = N.$$

On the other hand, let the random variable $\mathbf{y}$ be defined as

$$\mathbf{y} = \sum_{i=1}^{N} \widetilde{\mathbf{n}}_i^2$$

with $\widetilde{\mathbf{n}}_i$ being N mutually independent $N(0, \sigma_n)$ distributed random variables, then

[1] $u(x)$ denotes the unit step function.

$$\begin{aligned}\mathbf{y} &= \sum_{i=1}^{N} \widetilde{\mathbf{n}}_i^2 \\ &= \sum_{i=1}^{N} \left(\sigma_n \cdot \mathbf{n}_i\right)^2 \\ &= \sigma_n^2 \cdot \mathbf{x}\,.\end{aligned}$$

Hence, $\mathbf{y}$ has a PDF

$$f_{\mathbf{y}}(y) = \frac{1}{\sigma_n^2} \cdot f_{\mathbf{x}}\left(\frac{y}{\sigma_n^2}\right) \tag{C.1}$$

and a expected value

$$E\{\mathbf{y}\} = N \cdot \sigma_n^2.$$

Therefore, the r.v. $\mathbf{r}^2$ in Equation 3.11 can be generated by adding the square of 2 statistically independent mean free normal distributed r.v.s $\widetilde{\mathbf{n}}_1$ and $\widetilde{\mathbf{n}}_2$ with variance

$$\sigma_{\widetilde{n}} = \frac{1}{\sqrt{2}},$$

i.e.

$$\mathbf{r} = \widetilde{\mathbf{n}}_1^2 + \widetilde{\mathbf{n}}_2^2.$$

The energy amplification factor $\mathbf{r}^2$ has therefore a PDF of

$$f_{\mathbf{r}^2}(y) = \frac{x^{\frac{N}{2}-1}}{2^{\frac{N}{2}} \cdot \Gamma\left(\frac{N}{2}\right)} \cdot e^{-\frac{x}{2}} \cdot u(x)$$

$$\begin{aligned}f_{\mathbf{r}^2}(y) &= 2 \cdot f_{\mathbf{x}}(2 \cdot y)\,,\ N = 2 \\ &= e^{-y} \cdot u(y)\,.\end{aligned}$$

On the other hand, the signal amplification factor $\mathbf{r}$ in Figure 3.5 is simply obtained as the square root

$$\mathbf{r} = \sqrt{\widetilde{\mathbf{n}}_1^2 + \widetilde{\mathbf{n}}_1^2}$$

of the energy amplification.

Equivalently, the PDF of the r.v. $\mathbf{E}_S^{(Rx)}$ with expected value $\overline{E_S^{(T_x)}}$ can be obtained from Equation C.1 with

$$\begin{aligned}E\{\mathbf{y}\} &= 2 \cdot \sigma_n^2 \\ &= \overline{E_S^{(T_x)}}\end{aligned}$$

and hence $\sigma_n = \sqrt{\frac{\overline{E_S^{(T_x)}}}{2}}$. Substituting this value yields

$$\begin{aligned} f_{\mathbf{E}_S^{(Rx)}}(y) &= \frac{2}{\overline{E_S^{(T_x)}}} \cdot f_{\mathbf{x}}\left(\frac{2}{\overline{E_S^{(T_x)}}} \cdot y\right),\ N=2 \\ &= \frac{1}{\overline{E_S^{(T_x)}}} \cdot e^{-\frac{y}{\overline{E_S^{(T_x)}}}} \cdot u(y). \end{aligned}$$

q.e.d.

C.1.2 Derivation of the Logarithmic χ^2 - Distribution

Equation 3.13 gives the PDF and the CDF of the receipt signal energy in a Rayleigh fading channel. The performance curves of digital communication systems are usually plotted versus the signal-to-noise ratio in dB. We therefore transform these relations into logarithmic measures. The relation between the linear energy x and its logarithmic measure y in dB is

$$y = g(x) = 10 \cdot log_{10}(x).$$

Hence, the range $I_x = \{x \in \Re \,|x \geq 0\}$ is transformed into $I_y = \Re$ and since g is monoton increasing its inverse g^{-1} exists. Therefore [Sta94],

$$f_{\mathbf{y}}(y) = \frac{f_{\mathbf{x}}\left(g^{-1}(y)\right)}{g'\left(g^{-1}(y)\right)},\ y \in I_y$$

and

$$F_{\mathbf{y}}(y) = F_{\mathbf{x}}\left(g^{-1}(y)\right),\ y \in I_y.$$

With the first derivation

$$g'(x) = \frac{10}{ln(10) \cdot x},$$

the inverse

$$g^{-1}(y) = 10^{\frac{y}{10}},$$

and the logarithmic measure of the average value

$$m_{dB} = 10 \cdot log_{10}(m)$$

it follows

$$g'\left(g^{-1}(y)\right) = \frac{10^{\frac{10-y}{10}}}{ln10}$$

and finally

$$f_{\mathbf{E}_{S_{dB}}^{(R_x)}}(y) = ln10 \cdot 10^{\frac{y-m_{dB}-10}{10}} \cdot e^{-10^{\frac{y-m_{dB}}{10}}}, \quad y \in \Re$$

$$F_{\mathbf{E}_{S_{dB}}^{(R_x)}}(y) = 1 - e^{-10^{\frac{y-m_{dB}}{10}}}, \quad y \in \Re.$$

Plots of these both functions are depicted in Figure 3.7 on Page 32.

C.2 Multipath Rayleigh Channel Derivations

C.2.1 PDF and CDF for Constant Energy Ratio Profile

With a constant energy profile, the r.v. $\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}$ is the sum of $2 \cdot L$ squared mean free normal distributed r.v.s $\mathbf{X}_j$ with variance σ^2:

$$\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0} = \sum_{j=1}^{2 \cdot L} \mathbf{X}_j^2 .$$

Hence, the expected value is

$$\begin{aligned} E\left\{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}\right\} &= \textstyle\sum_{j=1}^{2 \cdot L} \mathbf{E}\left\{\mathbf{X}_j^2\right\} \\ &= \textstyle\sum_{j=1}^{2 \cdot L} VAR\left\{\mathbf{X}_j\right\} \\ &= 2 \cdot L \cdot \sigma^2 . \end{aligned}$$

If we want to avoid the distinction between the average receive and transmit ratios we set

$$\frac{\overline{E_S^{(Tx)}}}{N_0} = \frac{\overline{E_S^{(Rx)}}}{N_0}$$

and it follows

$$\sigma^2 = \frac{1}{2 \cdot L} \cdot \frac{\overline{E_S^{(Tx)}}}{N_0}.$$

Then

$$\begin{aligned} \frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0} &= \sigma^2 \cdot \textstyle\sum_{j=1}^{2 \cdot L} \left(\frac{1}{\sigma} \cdot \mathbf{X}_j\right)^2 \\ &= \sigma^2 \cdot \textstyle\sum_{j=1}^{2 \cdot L} \widetilde{\mathbf{X}}_j^2 \\ &= \sigma^2 \cdot \mathbf{X} , \end{aligned}$$

where $\widetilde{\mathbf{X}}_j$ are $N(0,1)$ distributed and hence $\mathbf{X}$ is χ^2 distributed with degree of freedom $2 \cdot L$. Accordingly, $\mathbf{X}$ has the PDF [Bro95]

$$f_{\mathbf{X}}(x) = \frac{1}{2^L \cdot \Gamma(L)} \cdot x^{L-1} \cdot e^{-\frac{x}{2}} \cdot u(x)$$

and $\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}$ has the scaled PDF

$$\begin{aligned} f_{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}}(x) &= \tfrac{1}{\sigma^2} \cdot f_{\mathbf{X}}\left(\tfrac{x}{\sigma^2}\right) \\ &= \tfrac{x^{L-1}}{\alpha^L \cdot \Gamma(L)} \cdot e^{-\frac{x}{\alpha}} \cdot u(x) \end{aligned}$$

with

$$\begin{aligned} \alpha &= 2 \cdot \sigma^2 \\ &= \tfrac{1}{L} \cdot \overline{\tfrac{E_S^{(Tx)}}{N_0}}. \end{aligned}$$

q.e.d.

For the CDF follows

$$\begin{aligned} F_{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}}(x) &= \int_{-\infty}^{x} f_{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}}(v)\, dv \\ &= \int_0^x \tfrac{v^{L-1}}{\alpha^L \cdot \Gamma(L)} \cdot e^{-\frac{v}{\alpha}} dv \cdot u(x). \end{aligned} \tag{C.2}$$

This integral is similar to the integral by which the incomplete Gamma function $\Gamma_{inc}(x,n)$ is defined [Abr65]:

$$\Gamma_{inc}(x,n) = \frac{1}{\Gamma(n)} \cdot \int_0^x v^{n-1} \cdot e^{-v} dv$$

In fact with the substitution $v = \alpha \cdot u$ in Equation C.2 we obtain

$$\begin{aligned} F_{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}}(x) &= \tfrac{1}{\Gamma(L)} \cdot \int_0^{\frac{x}{\alpha}} u^{L-1} \cdot e^{-u} du \\ &= \Gamma_{inc}\left(\tfrac{x}{\alpha}, L\right). \end{aligned} \tag{C.3}$$

Equation C.3 is defined for any $L \geq 1$, also for non-integer values. In our case, however, L takes only integer values $L = 1, 2, \ldots$ and in this case the integral can be recursively solved by partial integration and we obtain

$$\begin{aligned} F_{\frac{\mathbf{E}_S^{(Rx)}}{\mathbf{N}_0}}(x) &= \left[\tfrac{e^{-\frac{u}{\alpha}} \cdot (L-1)!}{\Gamma(L)} \cdot \left[\tfrac{u^{L-1}}{(L-1)!} + \tfrac{\alpha \cdot u^{L-2}}{(L-2)!} + \cdots + \alpha^{L-1}\right]\right]_0^x \\ &= \left[1 - e^{-\frac{x}{\alpha}} \cdot \textstyle\sum_{k=0}^{L-1} \tfrac{1}{k!} \cdot \left(\tfrac{x}{\alpha}\right)^k\right] \cdot u(x). \end{aligned}$$

q.e.d.

C.2.2 PDF and CDF for General Energy Ratio Profiles

A random variable created by the summation of the square of two independent $N(0,1)$ distributed random variables is said to be χ^2 distributed with degree of freedom 2. Its PDF is given by [Bro95]

$$f_{\chi^2}(x) = \begin{cases} \frac{1}{2} \cdot e^{-\frac{x}{2}}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

In the fist step we derive the PDF of a weighted sum of two squared $N(0,\sigma)$ distributed random variables. We therefore consider

$$\mathbf{z} = a \cdot \left(\mathbf{y}_1^2 + \mathbf{y}_2^2\right)$$

$$\mathbf{z} = a \cdot \mathbf{y}_1^2 + a \cdot \mathbf{y}_2^2$$

$$\mathbf{z} = \left(\sqrt{a}\mathbf{y}_1\right)^2 + \left(\sqrt{a}\mathbf{y}_2\right)^2$$

where the independent random variables $\mathbf{y}_1$ and $\mathbf{y}_2$ are both $N(0,\sigma)$ distributed and a is a positive scalar. Clearly, $\mathbf{z}$ is the sum of the square of two independent $N(0,\sqrt{a}\cdot\sigma)$ distributed random variables and hence obeys a weighted χ^2 distribution of degree of freedom 2:

$$f_{\mathbf{z}}(z) = \frac{1}{a \cdot \sigma^2} f_{\chi^2}\left(\frac{z}{a \cdot \sigma^2}\right)$$

$$f_{\mathbf{z}}(z) = \frac{1}{2 \cdot a \cdot \sigma^2} \cdot e^{-\frac{z}{2 \cdot a \cdot \sigma^2}} \cdot u(z)$$

$$f_{\mathbf{z}}(z) = \frac{1}{\alpha} e^{-\frac{z}{\alpha}} u(z) \tag{C.4}$$

with $\alpha = 2 \cdot a \cdot \sigma^2$ and $u(z)$ denoting the unit step function

$$u(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}.$$

Let us now define a new random variable $\mathbf{x}$ as the sum of n weighted χ^2 random variables of degree of freedom 2, all being mutually independent:

$$\mathbf{x} = \sum_{i=1}^{n} a_i \cdot \left(\mathbf{y}_{1,i}^2 + \mathbf{y}_{2,i}^2\right),\ n \geq 2$$

Again, $\mathbf{y}_{1,i}$ and $\mathbf{y}_{2,i}$ are $N(0,\sigma_i)$ distributed. For now, let us assume that all $\alpha_i = 2 \cdot a_i \cdot \sigma_i^2$ are different. Then, the PDF of $\mathbf{x}$ is given by

$$f_{\mathbf{x}}^{(n)}(x) = \sum_{i=1}^{n} \frac{a_i^{n-2} \cdot e^{-\frac{x}{a_i}}}{\prod_{j=1, j\neq i}^{n} (a_i - a_j)},\ n \geq 2. \tag{C.5}$$

This theorem if proofed in the following by complete induction. Equation C.5 yields for $n = 2$

$$f_{\mathbf{x}}^{(2)}(x) = \frac{1}{\alpha_1 - \alpha_2} \cdot e^{-\frac{x}{\alpha_1}} + \frac{1}{\alpha_2 - \alpha_1} \cdot e^{-\frac{x}{\alpha_2}} \tag{C.6}$$

Since $\mathbf{x}$ is the sum of two independent random variables, its PDF is obtained by the convolution of the two individual PDFs [Sta94]:

$$f_{\mathbf{x}}^{(2)}(x) = \int_{-\infty}^{\infty} \frac{1}{\alpha_1} e^{-\frac{v}{\alpha_1}} u(u) \cdot \frac{1}{\alpha_2} e^{-\frac{x-v}{\alpha_2}} u(x-v)\, dv$$

$$= \int_{0}^{x} \frac{1}{\alpha_1} e^{-\frac{v}{\alpha_1}} \cdot \frac{1}{\alpha_2} e^{-\frac{x-v}{\alpha_2}} dv$$

$$= \frac{e^{-\frac{x}{\alpha_2}}}{\alpha_1 \alpha_2} \int_{0}^{x} e^{\frac{\alpha_1 - \alpha_2}{\alpha_1 \alpha_2} v} dv$$

$$= \frac{e^{-\frac{x}{\alpha_2}}}{\alpha_1 - \alpha_2} \left[e^{\frac{\alpha_1 - \alpha_2}{\alpha_1 \alpha_2} x} - 1 \right]$$

$$= \frac{1}{\alpha_1 - \alpha_2} \left[e^{-\frac{x}{\alpha_1}} - e^{-\frac{x}{\alpha_2}} \right]$$

$$= \frac{1}{\alpha_1 - \alpha_2} e^{-\frac{x}{\alpha_1}} + \frac{1}{\alpha_2 - \alpha_1} e^{-\frac{x}{\alpha_2}}$$

Hence, Equation C.5 is valid for $n = 2$. Now we conclude from $f_{\mathbf{x}}^{(n)}$ to $f_{\mathbf{x}}^{(n+1)}$ by convolving $f_{\mathbf{x}}^{(n)}$ with the PDF of a weighted χ^2 r.v. (Equation C.4).

$$f_{\mathbf{x}}^{(n+1)}(x) = \int_{-\infty}^{\infty} f_{\mathbf{x}}^{(n)}(v) \cdot \frac{1}{\alpha_{n+1}} e^{-\frac{x-v}{\alpha_{n+1}}} u(x-v)\, dv$$

$$= e^{-\frac{x}{\alpha_{n+1}}} \int_{0}^{x} \left(\sum_{i=1}^{n} \frac{\alpha_i^{n-2} e^{-\frac{v}{\alpha_i}} e^{\frac{v}{\alpha_{n+1}}}}{\alpha_{n+1} \cdot \prod_{j=1, j\neq i}^{n} (\alpha_i - \alpha_j)} \right) dv$$

$$= e^{-\frac{x}{\alpha_{n+1}}} \sum_{i=1}^{n} \left(\frac{\alpha_i^{n-2}}{\alpha_{n+1} \prod_{j=1, j\neq i}^{n} (\alpha_i - \alpha_j)} \int_0^x e^{-\frac{\alpha_i - \alpha_{n+1}}{\alpha_i \alpha_{n+1}} v} dv \right)$$

$$= \sum_{i=1}^{n} \left[\frac{\alpha_i^{n-1} e^{-\frac{x}{\alpha_{n+1}}}}{\prod_{j=1, j\neq i}^{n+1} (\alpha_i - \alpha_j)} \left(e^{\frac{\alpha_i - \alpha_{n+1}}{\alpha_i \alpha_{n+1}} x} - 1 \right) \right]$$

$$= \sum_{i=1}^{n} \frac{\alpha_i^{n-1} e^{-\frac{x}{\alpha_i}}}{\prod_{j=1, j\neq i}^{n+1} (\alpha_i - \alpha_j)} - \sum_{i=1}^{n} \frac{\alpha_i^{n-1} e^{-\frac{x}{\alpha_{n+1}}}}{\prod_{i=1, i\neq j}^{n+1} (\alpha_i - \alpha_j)}$$

It will be shown that

$$\sum_{i=1}^{n} \frac{-\alpha_i^{n-1}}{\prod_{i=1, i\neq j}^{n+1} (\alpha_i - \alpha_j)} = \frac{\alpha_{n+1}^{n-1}}{\prod_{j=1}^{n} (\alpha_{n+1} - \alpha_j)} \tag{C.7}$$

Using this result

$$f_{\mathbf{x}}^{(n+1)}(x) = \sum_{i=1}^{n} \frac{\alpha_i^{n-1} e^{-\frac{x}{\alpha_i}}}{\prod_{j=1, j\neq i}^{n+1} (\alpha_i - \alpha_j)} + \frac{\alpha_{n+1}^{n-1}}{\prod_{j=1}^{n} (\alpha_{n+1} - \alpha_j)}$$

$$= \sum_{i=1}^{n} \frac{\alpha_i^{n-1} e^{-\frac{x}{\alpha_i}}}{\prod_{j=1, j\neq i}^{n+1} (\alpha_i - \alpha_j)} + \frac{\alpha_{n+1}^{n-1}}{\prod_{j=1}^{n} (\alpha_{n+1} - \alpha_j)}$$

$$= \sum_{i=1}^{n} \frac{\alpha_i^{n-1} e^{-\frac{x}{\alpha_i}}}{\prod_{j=1, j\neq i}^{n+1} (\alpha_i - \alpha_j)} + \frac{\alpha_{n+1}^{n-1}}{\prod_{j=1, j\neq n+1}^{n+1} (\alpha_{n+1} - \alpha_j)}$$

$$= \sum_{i=1}^{n+1} \frac{\alpha_i^{(n+1)-2} e^{-\frac{x}{\alpha_i}}}{\prod_{j=1, j\neq i}^{n+1} (\alpha_i - \alpha_j)}$$

q.e.d.

Proof of Equation C.7

The right side of Equation C.7 yields with $x = \alpha_{n+1}$

$$\frac{x^{n+1}}{\prod_{j=1}^{n} (x - \alpha_j)} = \frac{P(x)}{Q(x)}. \tag{C.8}$$

A rational function of this type with $Q(x)$ and $P(x)$ prime, $grad(Q(x)) > grad(P(x))$, and all α_j different and real can be decomposed into the following form [Bro95]

$$\frac{P(x)}{Q(x)} = \sum_{j=1}^{n} \frac{A_j}{x - \alpha_j}$$

with

$$A_j = \frac{P(\alpha_j)}{Q'(\alpha_j)}.$$

In our case

$$Q'(x) = \frac{d}{dx}\left[(x - \alpha_j) \cdot \prod_{i=1, i\neq j}^{n} (x - \alpha_i)\right]$$

$$= \prod_{i=1, i\neq j}^{n} (x - \alpha_i) + (x - \alpha_j) \cdot \frac{d}{dx} \prod_{i=1, i\neq j}^{n} (x - \alpha_i)$$

and hence

$$Q'(\alpha_j) = \prod_{i=1, i\neq j}^{n} (\alpha_j - \alpha_i)$$

$$A_j = \frac{\alpha_j^{n+1}}{\prod_{i=1, i\neq j}^{n} (\alpha_j - \alpha_i)}$$

and finally

$$\frac{x^{n+1}}{\prod_{j=1}^{n} (x - \alpha_j)} = \sum_{j=1}^{n} \frac{\alpha_j^{n+1}}{(x - \alpha_j) \prod_{i=1, i\neq j}^{n} (\alpha_j - \alpha_i)}$$

$$= \sum_{j=1}^{n} \frac{-\alpha_j^{n+1}}{(\alpha_j - x) \prod_{i=1, i\neq j}^{n} (\alpha_j - \alpha_i)}.$$

Reverse substituting $x = \alpha_{n+1}$ yields

$$\frac{\alpha_{n+1}^{n+1}}{\prod_{j=1}^{n} (\alpha_{n+1} - \alpha_j)} = \sum_{j=1}^{n} \frac{-\alpha_j^{n+1}}{(\alpha_j - \alpha_{n+1}) \prod_{i=1, i\neq j}^{n} (\alpha_j - \alpha_i)}$$

$$= \sum_{j=1}^{n} \frac{-\alpha_j^{n+1}}{\prod_{i=1, i\neq j}^{n+1} (\alpha_j - \alpha_i)}.$$

q.e.d.

C.3 Number of Transmissions Derivations

C.3.1 CDF of Memoryless CE-ARQ Systems

In Equation 4.10 the probability distribution $P(\mathbf{n}_{trans} = n),\ n \geq 1$ of the random variable $\mathbf{n}_{trans}$ is given. Hence, the cumulative distribution function is

$$\begin{aligned} P(\mathbf{n}_{trans} < x) &= 1 - P(\mathbf{n}_{trans} \geq x) \\ &= 1 - \textstyle\sum_{n=x}^{\infty} \left[P(R)^{n-1} \cdot (1 - P(R)) \right]. \end{aligned}$$

The term $1 - P(R)$ is independent of the summation index n and can be taken in front of the sum. After the index transformation we obtain

$$\begin{aligned} P(\mathbf{n}_{trans} < x) &= 1 - (1 - P(R)) \cdot \textstyle\sum_{n=0}^{\infty} P(R)^{x-1+n} \\ &= 1 - (1 - P(R)) \cdot P(R)^{x-1} \cdot \textstyle\sum_{n=0}^{\infty} P(R)^{n}. \end{aligned}$$

The remaining sum is the geometric series which is equal to $(1 - P(R))^{-1}$ and finally

$$P(\mathbf{n}_{trans} < x) = 1 - P(R)^{x-1}.$$

q.e.d.

C.3.2 CDF of MARQ Systems

In Equation 4.13 the probability distribution $P(\mathbf{n}_{trans} = n),\ n \geq 1$ of the random variable $\mathbf{n}_{trans}$ is given. Hence, the cumulative distribution function is

$$\begin{aligned} P(\mathbf{n}_{trans} < x) &= 1 - P(\mathbf{n}_{trans} \geq x) \\ &= 1 - \textstyle\sum_{n=x}^{\infty} \left[P\left(\bigcap_{j=1}^{n-1} R_j \right) - P\left(\bigcap_{j=1}^{n} R_j \right) \right] \\ &= 1 - \lim_{N \to \infty} \left\{ \textstyle\sum_{n=x}^{N} P\left(\bigcap_{j=1}^{n-1} R_j \right) - \sum_{n=x}^{N} P\left(\bigcap_{j=1}^{n} R_j \right) \right\}. \end{aligned}$$

Splitting the two sums and applying an index transformation yields

$$\begin{aligned} P(\mathbf{n}_{trans} < x) &= 1 - \lim_{N \to \infty} \left\{ P\left(\bigcap_{j=1}^{x-1} R_j \right) + \textstyle\sum_{n=x}^{N-1} P\left(\bigcap_{j=1}^{n} R_j \right) \right. \\ &\qquad \left. - \textstyle\sum_{n=x}^{N-1} P\left(\bigcap_{j=1}^{n} R_j \right) - P\left(\bigcap_{j=1}^{N} R_j \right) \right\} \\ &= 1 - \lim_{N \to \infty} \left\{ P\left(\bigcap_{j=1}^{x-1} R_j \right) - P\left(\bigcap_{j=1}^{N} R_j \right) \right\}. \end{aligned}$$

Consequently,

$$P\left(\mathbf{n}_{trans} < x\right) = 1 - P\left(\bigcap_{j=1}^{x-1} R_j\right)$$

if

$$\lim_{N\to\infty}\left\{P\left(\bigcap_{j=1}^{N} R_j\right)\right\} = 0. \tag{C.9}$$

Equation C.9 is in general not true. As an example consider a (useless) ARQ system which does not attempt to decode after n transmissions. Then

$$\lim_{N\to\infty}\left\{P\left(\bigcap_{j=1}^{N} R_j\right)\right\} = P\left(\bigcap_{j=1}^{n} R_j\right).$$

However, all reasonable ARQ systems, independent of their type, have at least a constant error correction capability. That is, the smallest error correction capability is already achieved after the first transmission and each additional transmission has the same (memoryless CE-ARQ) or an increased error correction capability (MARQ). For memoryless CE-ARQ systems, the probability of the intersection event in Equation C.9 is identical to $P\left(R\right)^N$ and the proof is straight forward. MARQ systems, however, deliver additional redundancy. The worst case to be considered is when the redundant information is delivered in the smallest possible increments. The crucial point is that even if the increments are extremely small it takes only a *finite* (but possibly extremely high) number of transmissions L to deliver that much additional information to the sink so that the MARQ system outperforms the corresponding memoryless scheme having 2 transmissions available, i.e.

$$P\left(\bigcap_{i=1}^{L+1} R_i\right) \leq P\left(R_1\right)^2.$$

After L additional transmission the system also outperforms the correspondingmemoryless scheme with 3 transmissions, and so on. Consequently,

$$P\left(\bigcap_{i=1}^{N} R_i\right) \leq P\left(R_1\right)^{\lfloor\frac{N-1}{L}\rfloor+1}$$

and the limes in Equation C.9 can be bounded as

$$0 \leq \lim_{N\to\infty}\left\{P\left(\bigcap_{i=1}^{N} R_i\right)\right\} \leq \lim_{N\to\infty}\left\{P\left(R_1\right)^{\lfloor\frac{N-1}{L}\rfloor+1}\right\} = 0 \tag{C.10}$$

which proofs Equation C.9.

q.e.d.

C.3.3 $\overline{n_{trans}}$ of Memoryless CE-ARQ Systems

In Equation 4.10 the probability distribution $P(\mathbf{n}_{trans} = n),\ n \geq 1$ of the random variable $\mathbf{n}_{trans}$ is given. The expected value of this discrete distribution is

$$\begin{aligned}
\overline{n_{trans}} &= E\{\mathbf{n}_{trans}\} = \textstyle\sum_{n=1}^{\infty} n \cdot P(\mathbf{n}_{trans} = n) \\
&= \textstyle\sum_{n=1}^{\infty} n \cdot P(R)^{n-1} \cdot (1 - P(R)) \\
&= \lim_{N\to\infty} \left(\textstyle\sum_{n=1}^{N} n \cdot P(R)^{n-1} - \sum_{n=1}^{N} n \cdot P(R)^{n} \right) \\
&= \lim_{N\to\infty} \left(1 + \textstyle\sum_{n=2}^{N} n \cdot P(R)^{n-1} - \sum_{n=2}^{N} (n-1) \cdot P(R)^{n-1} - N \cdot P(R)^{N} \right) \\
&= \lim_{N\to\infty} \left(1 + \textstyle\sum_{n=2}^{N} P(R)^{n-1} - N \cdot P(R)^{N} \right) \\
&= \lim_{N\to\infty} \left(\textstyle\sum_{n=0}^{N-1} P(R)^{n} - N \cdot P(R)^{N} \right)
\end{aligned}$$

With $0 < P(R) < 1$ it follows

$$\lim_{N\to\infty} \left(N \cdot P(R)^{N} \right) = 0 \tag{C.11}$$

and (geometric series)

$$\overline{n_{trans}} = \sum_{n=0}^{\infty} P(R)^{n} = \frac{1}{1 - P(R)}$$

q.e.d.

C.3.4 $\overline{n_{trans}}$ VE-ARQ and MARQ

Using probability distribution of $\mathbf{n}_{trans}$ in Equation 4.13, the expected value can be computed as

$$\begin{aligned}
E\{\mathbf{n}_{trans}\} &= \textstyle\sum_{n=1}^{\infty} n \cdot P(\mathbf{n}_{trans} = n) \\
&= 1 - P(R_1) + \textstyle\sum_{n=2}^{\infty} n \cdot P\left(\bigcap_{i=1}^{n-1} R_i \right) \cdot \left[1 - P\left(\bigcap_{i=1}^{n} R_i \,\middle|\, \bigcap_{i=1}^{n-1} R_i \right) \right] \\
&= 1 + \lim_{N\to\infty} \left\{ \textstyle\sum_{n=2}^{N} n \cdot P\left(\bigcap_{i=1}^{n-1} R_i \right) - \sum_{n=1}^{N} n \cdot P\left(\bigcap_{i=1}^{n} R_i \right) \right\} \\
&= 1 + \lim_{N\to\infty} \left\{ \textstyle\sum_{n=1}^{N-1} (n+1) \cdot P\left(\bigcap_{i=1}^{n} R_i \right) - \sum_{n=1}^{N} n \cdot P\left(\bigcap_{i=1}^{n} R_i \right) \right\} \\
&= 1 + \lim_{N\to\infty} \left\{ \textstyle\sum_{n=1}^{N-1} P\left(\bigcap_{i=1}^{n} R_i \right) - N \cdot P\left(\bigcap_{i=1}^{N} R_i \right) \right\}
\end{aligned}$$

and hence

$$E\left\{\mathbf{n}_{trans}\right\} = 1 + \sum_{n=1}^{\infty} P\left(\bigcap_{i=1}^{n} R_i\right),$$

if

$$\lim_{N\to\infty}\left\{N \cdot P\left(\bigcap_{i=1}^{N} R_i\right)\right\} = 0 \tag{C.12}$$

which remains to be shown. In Section C.3.2, we already dealt with the similar limit

$$\lim_{N\to\infty}\left\{P\left(\bigcap_{i=1}^{N} R_i\right)\right\}$$

which was shown to be zero for usefull MARQ systems by bounding it with the rejection probability of a memoryless CE-ARQ system (Inequation C.10). Applying the same argumentation yields the inequation

$$0 \leq \lim_{N\to\infty}\left\{N \cdot P\left(\bigcap_{i=1}^{N} R_i\right)\right\} \leq \lim_{N\to\infty}\left\{N \cdot P\left(R_1\right)^{\left\lfloor\frac{N-1}{L}\right\rfloor+1}\right\}.$$

Since in general

$$\lim_{x\to\infty}\left\{x \cdot K^x\right\} = 0,\ 0 \leq K < 1$$

it follows

$$\lim_{N\to\infty}\left\{N \cdot P\left(\bigcap_{i=1}^{N} R_i\right)\right\} = 0$$

q.e.d.

C.4 Throughput Derivations

C.4.1 AWGN Channel Capacities versus $\frac{E_b}{N_0}$

In Section 4.4.3 it was mentioned that the channel capacities for the AWGN can be represented as function of the normalized bit energy to the noise power density $\frac{E_b}{N_0}$. Therefore, it needs to be proven that the fixed point Equation 4.25

$$R_{max} = C\left(R_{max}\,\frac{symbol}{bit} \cdot \frac{E_b}{N_0}\right)$$

has one unique fixed point R_{max} for each $\frac{E_b}{N_0}$.

The channel capacity for the 1-dimensional AWGN was given in Equation 4.22 as

$$C_{1D-AWGN}\left(x\right) = \frac{1}{2} \cdot log_2\left(1 + 2 \cdot x\right)\frac{bit}{symbol},\ x \geq 0.$$

Hence, the right side of the fixed point Equation 4.25 is

$$C_{1D-AWGN}\left(R_{max}, \frac{E_b}{N_0}\right) = \frac{1}{2} \cdot log_2\left(1 + 2 \cdot R_{max} \cdot \frac{E_b}{N_0} \cdot \frac{symbol}{bit}\right) \frac{bit}{symbol}, R_{max}, \frac{E_b}{N_0} \geq 0.$$

From that we derive that

$$C_{1D-AWGN}\left(R_{max}, \frac{E_b}{N_0}\right) \geq 0 \; \forall \, R_{max}, \frac{E_b}{N_0} \geq 0$$

with

$$C_{1D-AWGN}\left(0, \frac{E_b}{N_0}\right) = 0 \; \forall \, \frac{E_b}{N_0} \geq 0. \tag{C.13}$$

Hence $R_{max} = 0$ is already a fixed point for all $\frac{E_b}{N_0} \geq 0$. In order to search for non-trivial fixed points at $R_{max} > 0$ we obtain the the right-sided derivative as

$$\frac{d}{dR_{max}} C_{1D-AWGN}\left(R_{max}, \frac{E_b}{N_0}\right) = \frac{1}{ln\,(2)} \frac{\frac{E_b}{N_0}}{1 + 2 \cdot R_{max} \cdot \frac{E_b}{N_0} \cdot \frac{symbol}{bit}}, R_{max}, \frac{E_b}{N_0} \geq 0. \tag{C.14}$$

Clearly, in order to obtain further fixed points we need to restrict $\frac{E_b}{N_0} > 0$. Then it follows that

$$\begin{aligned} \max_{R_{max}} \left\{ \frac{d}{dR_{max}} C_{1D-AWGN}\left(R_{max}, \frac{E_b}{N_0}\right) \right\} &= \frac{d}{dR_{max}} C_{1D-AWGN}\left(R_{max}, \frac{E_b}{N_0}\right) \\ &= \frac{1}{ln(2)} \cdot \frac{E_b}{N_0}. \end{aligned}$$

and that the derivative steadily decreases and approaches its limit value 0 for $R_{max} \to \infty$. Since the derivative of the left side of 4.25 is 1, we distinguish the two cases $\frac{E_b}{N_0} > ln\,(2)$ and $\frac{E_b}{N_0} \leq ln\,(2)$. In the first case, the slope of $C_{1D-AWGN}\left(R_{max}, \frac{E_b}{N_0}\right)$ is larger than 1 for all $\frac{E_b}{N_0} > 0$ and we have one additional fixed point at some value $R_{max} > 0$. In the latter case there is no further fixed point. Hence, the fixed point Equation 4.25 has one unique fixed point for $R_{max} > 0$ and $\frac{E_b}{N_0} > ln\,(2) = -1.59\,dB$. Within that ranges, the channel capacity $C = R_{max}$ of the 1-dim AWGN channel can be represented as a function of $\frac{E_b}{N_0}$.

For the 2-dimensional case the same argumentation holds. Again, $R_{max} = 0$ is a trivial fixed point and the derivative $\frac{d}{dR_{max}} C_{2D-AWGN}\left(R_{max}, \frac{E_b}{N_0}\right)$ has the same functional properties, especially

$$\max_{R_{max}} \left\{ \frac{d}{dR_{max}} C_{1D-AWGN}\left(R_{max}, \frac{E_b}{N_0}\right) \right\} = \frac{1}{ln\,(2)} \cdot \frac{E_b}{N_0}.$$

Hence, the channel capacity of the 2-dimensional AWGN channel can be represented as a function of $\frac{E_b}{N_0}$for $C > 0$ and $\frac{E_b}{N_0} > -1.59\,dB$.

C.4.2 Expected Value of Total Number of Required Symbols $\mathbf{L}_{tot}$

In the following Equation 4.32 is proofed. Using the probability distribution of $\mathbf{L}_{tot}$ given in Equation 4.31, its expected value is

$$\begin{aligned} E\{\mathbf{L}_{tot}\} &= L_1 \cdot P(\mathbf{n}_{trans}=1) + (L_1+L_2)\cdot P(\mathbf{n}_{trans}=2) + \cdots \\ &= \textstyle\sum_{n=1}^{\infty}\left[\left(\sum_{j=1}^{n} L_j\right)\cdot P(\mathbf{n}_{trans}=n)\right]. \end{aligned} \tag{C.15}$$

Substituting the probability distribution Equation 4.13 yields

$$\begin{aligned} E\{\mathbf{L}_{tot}\} &= L_1\cdot(1-P(R_1)) \\ &\quad + (L_1+L_2)\cdot P(R_1)\cdot[1-P(R_1\cap R_2\,|R_1)] + \cdots \\ &\quad + \textstyle(\sum_{i=1}^{n} L_i)\cdot P\left(\bigcap_{i=1}^{n-1} R_i\right)\cdot\left[1-P\left(\bigcap_{i=1}^{n} R_i\,\middle|\bigcap_{i=1}^{n-1} R_i\right)\right] + \cdots \\ &= L_1\cdot(1-P(R_1)) \\ &\quad + \textstyle\sum_{n=2}^{\infty}\left[(\sum_{i=1}^{n} L_i)\cdot P\left(\bigcap_{i=1}^{n-1} R_i\right)\cdot\left[1-P\left(\bigcap_{i=1}^{n} R_i\,\middle|\bigcap_{i=1}^{n-1} R_i\right)\right]\right]. \end{aligned}$$

Now, the infinite sum can be expressed as the limes of a finite sum:

$$\begin{aligned} E\{\mathbf{L}_{tot}\} &= L_1\cdot(1-P(R_1)) + \\ &\quad \lim_{N\to\infty}\left\{\textstyle\sum_{n=2}^{N}\left[(\sum_{i=1}^{n} L_i)\cdot P\left(\bigcap_{i=1}^{n-1} R_i\right) - (\sum_{i=1}^{n} L_i)\cdot P\left(\bigcap_{i=1}^{n} R_i\right)\right]\right\} \\ &= L_1\cdot(1-P(R_1)) + \lim_{N\to\infty}\left\{\textstyle\sum_{n=2}^{N} L_n\cdot P\left(\bigcap_{i=1}^{n-1} R_i\right)\right. \\ &\quad + \textstyle\sum_{n=2}^{N}\left[\left(\sum_{i=1}^{n-1} L_i\right)\cdot P\left(\bigcap_{i=1}^{n-1} R_i\right)\right] \\ &\quad \left. - \textstyle\sum_{n=2}^{N}\left[(\sum_{i=1}^{n} L_i)\cdot P\left(\bigcap_{i=1}^{n} R_i\right)\right]\right\} \\ &= L_1 + \lim_{N\to\infty}\left\{\textstyle\sum_{n=2}^{N} L_n\cdot P\left(\bigcap_{i=1}^{n-1} R_i\right)\right. \\ &\quad + \textstyle\sum_{n=1}^{N-1}\left[(\sum_{i=1}^{n} L_i)\cdot P\left(\bigcap_{i=1}^{n} R_i\right)\right] \\ &\quad \left. - \textstyle\sum_{n=2}^{N}\left[(\sum_{i=1}^{n} L_i)\cdot P\left(\bigcap_{i=1}^{n} R_i\right)\right]\right\} \\ &= L_1 + \lim_{N\to\infty}\left\{\textstyle\sum_{n=2}^{N} L_n\cdot P\left(\bigcap_{i=1}^{n-1} R_i\right)\right. \\ &\quad \left. - \left(\textstyle\sum_{i=1}^{N} L_i\right)\cdot P\left(\bigcap_{i=1}^{N} R_i\right)\right\} \end{aligned}$$

Hence,

$$E\{\mathbf{L}_{tot}\} = L_1 + \sum_{n=2}^{\infty} L_n\cdot P\left(\bigcap_{i=1}^{n-1} R_i\right)$$

if

$$\lim_{N\to\infty}\left\{\left(\sum_{i=1}^{N} L_i\right)\cdot P\left(\bigcap_{i=1}^{N} R_i\right)\right\} = 0,$$

which will be shown in the following: All packet sizes L_n are finite. With

$$L_{max} = \max_{n,\, n \geq 1} \{L_n\}$$

it follows

$$\sum_{i=1}^{N} L_i \leq N \cdot L_{max}$$

and

$$0 \leq \lim_{N \to \infty} \left\{ \left(\sum_{i=1}^{N} L_i \right) \cdot P \left(\bigcap_{i=1}^{N} R_i \right) \right\} \leq L_{max} \cdot \lim_{N \to \infty} \left\{ N \cdot P \left(\bigcap_{i=1}^{N} R_i \right) \right\}.$$

In Section C.3.4 is was shown that

$$\lim_{N \to \infty} \left\{ N \cdot P \left(\bigcap_{i=1}^{N} R_i \right) \right\} = 0$$

and hence,

$$0 \leq \lim_{N \to \infty} \left\{ \left(\sum_{i=1}^{N} L_i \right) \cdot P \left(\bigcap_{i=1}^{N} R_i \right) \right\} \leq 0.$$

q.e.d.

C.5 Delay Derivations

C.5.1 Derivation of Bounds for Average Packet Delays

Equations 4.40 - 4.43: Stop-and-Wait Protocol

In Equation 4.39, the probability distribution of the average packet delay $\Delta \mathbf{t}_p$ was given. Hence, its expected value is

$$\begin{aligned} \overline{\Delta t_p} &= E\{\Delta \mathbf{t}_p\} \\ &= \sum_{n=1}^{\infty} \left[\left(\sum_{j=1}^{n} t_j + n \cdot t_{RT} \right) \cdot P\left(\mathbf{n}_{trans} = n\right) \right]. \end{aligned}$$

Using Equation 4.35, $\overline{\Delta t_p}$ can be expressed by the expected value of the total transmitted symbols and the average number of transmissions:

$$\begin{aligned}\overline{\Delta t_p} &= t_S \cdot \sum_{n=1}^{\infty} \left[\left(\sum_{j=1}^{n} L_j \right) \cdot P\left(\mathbf{n}_{trans} = n\right) \right] \\ &\quad + t_{RT} \cdot \sum_{n=1}^{\infty} \left[n \cdot P\left(\mathbf{n}_{trans} = n\right) \right] \\ &= t_S \cdot E\left\{\mathbf{L}_{tot}\right\} + t_{RT} \cdot \overline{n_{trans}}\end{aligned} \tag{C.16}$$

After applying the corresponding Bounds 4.19 and 4.33 to Equation C.16 we proof Bounds 4.40 to 4.42

$$\begin{aligned} & t_{RT} + L_1 \cdot t_S + \sum_{n=1}^{\infty} \left[\left(t_{RT} + L_{n+1} \cdot t_S \right) \cdot \prod_{i=1}^{n} P\left(R_i\right) \right] \\ \leq\ & \overline{\Delta t_p} \\ \leq\ & t_{RT} + L_1 \cdot t_S + \sum_{n=1}^{\infty} \left[\left(t_{RT} + L_{n+1} \cdot t_S \right) \cdot P\left(R_n\right) \right] \end{aligned}$$

and substitution of Equation 4.20 in Equation C.16 yields

$$\overline{\Delta t_p} = t_{RT} \cdot \overline{n_{trans}} + t_S \cdot \frac{L_{info}}{T}.$$

Normalization with the round trip delay t_{RT} and application of Equation 4.37 finally leads to Equation 4.43:

$$\begin{aligned}\frac{\overline{\Delta t_p}}{t_{RT}} &= \overline{n_{trans}} + \frac{t_S \cdot L_{info}}{t_{RT}} \cdot \frac{1}{T} \\ &= \overline{n_{trans}} + \frac{1}{N_{RT}} \cdot \frac{1}{T}.\end{aligned}$$

q.e.d.

Go-Back-N Protocol

C.5.2 Derivation of Bounds for Average Information Frame Delay

Go-Back-N Protocol

Selective-Repeat Protocol

As mentioned the actual distribution as well as the expected value of these delay contribution are hard to derive. In the following we will derive the minimum round trip delay contribution and under which circumstances it can be attained. If in time step 5 the packet #2 would have been positively acknowledged and also the three remaining packets 3-5 in time step 7-9, this would

result in the smallest possible round trip delay

$$\left.\frac{\overline{\Delta t}_I^{(N_I)}}{t_{RT}}\right|_{round\,trip\,MIN} = 1.$$

However, already the probability of that event is hard to derive: Clearly, if the N_{RT} last packets have not been previously negatively acknowledged, the probability of interest is $[P(\mathbf{n}_{trans}=1)]^{N_{RT}}$. But suppose packet #1 in time step 5 would have been negatively acknowledged. Then, the last 4 packets consist of of 3 packets which are transmitted for the first time and one packet transmitted for the second time. Hence, the probability of the shortest round trip delay in that case is $[P(\mathbf{n}_{trans}=1)]^{N_{RT}-1}\cdot P(\mathbf{n}_{trans}=2)$. Furthermore, there are an infinite number of compositions of the last N_{RT} packets of an information frame and their distribution is not obtainable without a deeper journey into statistics. Nevertheless, we consider special cases where this minimum round trip delay contribution to the overall information frame delay as achieved. An obvious case is for high signal-to-noise ratios, i.e. where $P(\mathbf{n}_{trans}=1)\approx 1$ and $P(\mathbf{n}_{trans}=k)\approx 0,\ k\geq 2$. The corresponding round trip contribution of the SW protocol is N_I and for the GBN protocol also 1, so that the SR protocol does not provide a gain over the GBN protocol for that particular case. Now we consider a case which is at first not that obvious. Assume we use the ARQ system with the rejection probabilities depicted in Figure 4.5 on Page 58 and the average number of transmission shown in Figure 4.13 on Page 71 and we are operating at a signal-to-noise ratio range of $0\,dB < \frac{E_S}{N_0} < 2\,dB$. In that range $P(\mathbf{n}_{trans}=2)\approx 1$ and $P(\mathbf{n}_{trans}=k)\approx 0,\ k\neq 2$, that is we have almost exactly $\mathbf{n}_{trans}=2$ transmissions per packet. What happens in that case is illustrated in Figure . From time step 0 to 4, the round trip route fills with packets and in time step 5 the first packet is completely received. Since $P(\mathbf{n}_{trans}=1)\approx 0$, this packet is reject. Also packet #2 in time step 2 and so on. At time step 10 we have the same distribution of packets as in time step 5, but packets 1 to 5 have already been rejected once. In the following 5 time steps these packets are all accepted by the receiver since $P(\mathbf{n}_{trans}=2)\approx 1$. The same procedure occurs with packets 6 to 10 in the time steps 15 to 24. All are rejected once and finally accepted. In time step 25 the last remaining packet 11 is rejected. By now, the round trip contribution to the information frame delay is exactly the round trip delay. Packet 11 is retransmitted which adds again the round trip delay to the overall contribution.

To find a rule we generalize this special case. The first thing we notice is that if the information frame length is increased in multiples of $N_{RT}+1$ no extra time is added to the round trip delay contribution. Also the actual number N_{RT} is of no concern. If packet #11 would not be existent, the round trip delay contribution would only be the round trip delay and if one additional packet would be existent the round trip contribution would be smaller than $2\cdot t_{TR}$. Hence,

$$1 \leq \left.\frac{\overline{\Delta t}_I^{(N_I)}}{t_{RT}}\right|_{round\,trip} \leq 2,$$

for the given conditions $P(\mathbf{n}_{trans}=2)\approx 1$ and $P(\mathbf{n}_{trans}=k)\approx 0,\ k\neq 2$. If the n-th transmission has the probability one of being accepted and all other probabilities are zero, then the last packet

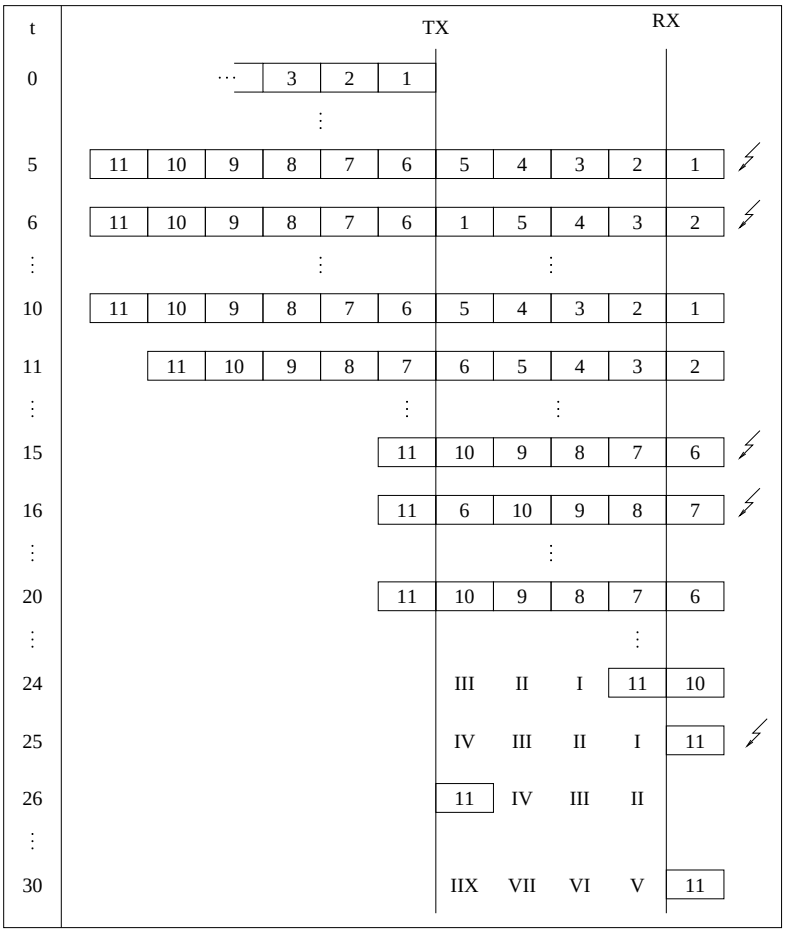

Figure C.1: Illustration

in the frame adds $(n-1)\cdot t_{RT}$ so that we obtain the general rule

$$1 \leq \left.\frac{\overline{\Delta t_I}^{(N_I)}}{t_{RT}}\right|_{round\ trip} \leq n,\ if\ P\left(\mathbf{n}_{trans}=n\right)\approx 1\ \wedge\ P\left(\mathbf{n}_{trans}=k\right)\approx 0,\ k\neq n. \tag{C.17}$$

C.6 Data Rate Derivations

C.6.1 Derivations Maximal Data Rate

The maximum data rate for the one dimensional AWGN channel was given in Equation 4.58. Hence,

$$\lim_{t_S\to 0^+}\left\{R_{Data}^{max,1D}\right\} = \lim_{t_S\to 0^+}\left\{\frac{\log_2\left(1+2\cdot\frac{P_S}{N_0}\cdot t_S\right)}{2\cdot t_S}\right\}$$

and application of L'Hospitals rule yields

$$\begin{aligned}\lim_{t_S\to 0^+}\left\{R_{Data}^{max,1D}\right\} &= \lim_{t_S\to 0^+}\left\{\frac{2\cdot\frac{P_S}{N_0}}{ln(2)\cdot\left(1+2\cdot\frac{P_S}{N_0}\cdot t_S\right)}\cdot\frac{1}{2}\right\}\\ &= \frac{1}{\ln(2)}\cdot\frac{P_S}{N_0}.\end{aligned}$$

Analog for the two dimensional AWGN channel, using Equation 4.59 for the limit and applying L'Hospitals rule yields

$$\begin{aligned}\lim_{t_S\to 0^+}\left\{R_{Data}^{max,2D}\right\} &= \lim_{t_S\to 0^+}\left\{\frac{\log_2\left(1+\frac{P_S}{N_0}\cdot t_S\right)}{t_S}\right\}\\ &= \lim_{t_S\to 0^+}\left\{\frac{\frac{P_S}{N_0}}{ln(2)\cdot\left(1+\frac{P_S}{N_0}\cdot t_S\right)}\cdot 1\right\}\\ &= \frac{1}{\ln(2)}\cdot\frac{P_S}{N_0}.\end{aligned}$$

q.e.d.

C.6.2 Limit of K^{SW} for $\frac{E_S}{N_0}\to-\infty$

For $\frac{E_S}{N_0}\to-\infty$, the number of transmission approaches infinity and the throughput approaches zero. Hence, to find the limit

$$\lim_{\frac{E_S}{N_0}\to-\infty}\left\{K^{SW}\right\} = \lim_{\frac{E_S}{N_0}\to-\infty}\left\{\frac{1}{N_{RT}\cdot\overline{n_{trans}}\cdot T+1}\right\}$$

we have to consider the limit of number transmissions times the corresponding throughput:

$$\lim_{n\to\infty}\left\{n\cdot R_{CM}^{ARQ,n}\right\}.$$

If the packet size is constant from th $\widetilde{n}$-th transmission, the coding-/modulation rate can written for $j \geq \widetilde{n}$ as (see Equation 4.3)

$$\begin{aligned} R_{CM}^{ARQ,n} &= \frac{1}{\sum_{i=1}^{\widetilde{n-1}} \frac{1}{R_{CM}^{FEC,i}} + \sum_{i=n}^{\widetilde{n}} \frac{1}{R_{CM}^{FEC,i}}} \\ &= \frac{1}{\sum_{i=1}^{\widetilde{n-1}} \frac{1}{R_{CM}^{FEC,i}} + \sum_{i=n}^{n} \widetilde{\frac{1}{R_{CM}^{FEC,n}}}} \\ &= \frac{1}{\sum_{i=1}^{\widetilde{n-1}} \frac{1}{R_{CM}^{FEC,i}} + \left(n+1-\widetilde{n}\right)\cdot \widetilde{\frac{1}{R_{CM}^{FEC,n}}}}. \end{aligned}$$

Hence, the limit can be obtained to

$$\begin{aligned} \lim_{n\to\infty}\left\{n\cdot R_{CM}^{ARQ,n}\right\} &= \lim_{n\to\infty}\left\{\frac{n}{\sum_{i=1}^{\widetilde{n-1}} \frac{1}{R_{CM}^{FEC,i}} + \left(n+1-\widetilde{n}\right)\cdot \widetilde{\frac{1}{R_{CM}^{FEC,n}}}}\right\} \\ &= \lim_{n\to\infty}\left\{\frac{1}{\frac{1}{n}\cdot\sum_{i=1}^{\widetilde{n-1}} \frac{1}{R_{CM}^{FEC,i}} + \left(1+\frac{\widetilde{1-n}}{n}\right)\cdot \widetilde{\frac{1}{R_{CM}^{FEC,n}}}}\right\} \\ &= R_{CM}^{FEC,\widetilde{n}} \end{aligned}$$

and finally

$$\begin{aligned} \lim_{\frac{E_S}{N_0}\to-\infty}\left\{K^{SW}\right\} &= \frac{1}{N_{RT}\cdot R_{CM}^{FEC,\widetilde{n}}+1} \\ &= \frac{1}{n_{RT}^{\left(\widetilde{n}\right)}+1}. \end{aligned}$$

q.e.d.

C.7 ML Combining Derivations

In the following, Equation 5.39 is proofed. Let c_j, $j \geq 0$ be real scalars with $c_j > 0$ and $\overrightarrow{a_j}$, $j \geq 0$ as well as $\overrightarrow{b}$ be arbitrary real vectors. Then,

$$
\begin{aligned}
\sum_{j=0}^{N-1} c_j \cdot \left|\overrightarrow{a_j} - \overrightarrow{b}\right|^2 &= \sum_{j=0}^{N-1} c_j \cdot \left(\left|\overrightarrow{a_j}\right|^2 - 2 \cdot \overrightarrow{a} \cdot \overrightarrow{b} + \left|\overrightarrow{b}\right|^2\right) \\
&= \sum_{j=0}^{N-1} c_j \cdot \left|\overrightarrow{a_j}\right|^2 - 2 \cdot \left(\sum_{j=0}^{N-1} c_j \cdot \overrightarrow{a_j}\right) \cdot \overrightarrow{b} + \left(\sum_{j=0}^{N-1} c_j\right) \cdot \left|\overrightarrow{b}\right|^2 \\
&= \sum_{j=0}^{N-1} c_j \cdot \left|\overrightarrow{a_j}\right|^2 + \left(\sum_{j=0}^{N-1} c_j\right) \cdot \left[-2 \cdot \frac{\sum_{j=0}^{N-1} c_j \cdot \overrightarrow{a_j}}{\sum_{j=0}^{N-1} c_j} + \left|\overrightarrow{b}\right|^2\right] \\
&= \sum_{j=0}^{N-1} c_j \cdot \left|\overrightarrow{a_j}\right|^2 - \frac{\left|\sum_{j=0}^{N-1} c_j \cdot \overrightarrow{a_j}\right|^2}{\sum_{j=0}^{N-1} c_j} + \left(\sum_{j=0}^{N-1} c_j\right) \cdot \left|\frac{\sum_{j=0}^{N-1} c_j \cdot \overrightarrow{a_j}}{\sum_{j=0}^{N-1} c_j} - \overrightarrow{b}\right|^2 \\
&= \frac{1}{\left(\sum_{j=0}^{N-1} c_j\right)} \cdot \left|\sum_{j=0}^{N-1} c_j \cdot \overrightarrow{a_j} - \left(\sum_{j=0}^{N-1} c_j\right) \cdot \overrightarrow{b}\right|^2 \\
&\qquad + f\left(\overrightarrow{a}_0, \ldots, \overrightarrow{a}_{N-1}, c_0, \ldots, c_{N-1}\right),
\end{aligned}
$$

where $f\left(\overrightarrow{a}_0, \ldots, \overrightarrow{a}_{N-1}, c_0, \ldots, c_{N-1}\right)$ is a scalar function independent of the vector $\overrightarrow{b}$.

q.e.d.

C.8 MAP Combining Derivations

C.8.1 Bayes Theorem for Mixed Random Variables

Let $\mathbf{x}$ be a discrete random variable with the sample space $\underline{x} = \{x_0, x_1, \ldots, x_{N-1}\}$ and with probabilities $P_{\mathbf{x}}(x)$, $x \in \underline{x}$. Furthermore, let $\mathbf{y}$ be another discrete random variable. The law of total probability states that

$$
P_{\mathbf{y}}(y) = \sum_{i=0}^{N-1} P_{\mathbf{y}|\mathbf{x}}(y|x_i) \cdot P_{\mathbf{x}}(x_i) \tag{C.18}
$$

where $P_{\mathbf{y}|\mathbf{x}}(y|x)$ denotes the conditioned probability of the event $\mathbf{y} = y$ under the assumption that $\mathbf{x} = x$ is observed. With the definition of the conditioned probability and Equation C.18 it follows Bayes Theorem for discrete random variables

$$
\begin{aligned}
P_{\mathbf{x}|\mathbf{y}}(x_i|y) &= \frac{P_{\mathbf{x},\mathbf{y}}(x_i,y)}{P_{\mathbf{y}}(y)} \\
&= \frac{P_{\mathbf{y}|\mathbf{x}}(y|x_i) \cdot P_{\mathbf{x}}(x_i)}{P_{\mathbf{y}}(y)} \\
&= \frac{P_{\mathbf{y}|\mathbf{x}}(y|x_i) \cdot P_{\mathbf{x}}(x_i)}{\sum_{i=0}^{N-1} P_{\mathbf{y}|\mathbf{x}}(y|x_i) \cdot P_{\mathbf{x}}(x_i)}
\end{aligned}
\quad .
$$

Equivalently, let $\mathbf{x}$ be a continuous random variable with the sample space $\underline{x}$ and with probability density function $f_{\mathbf{x}}(x)$ and let $\mathbf{y}$ be another continuous random variable. The total law of

probability states

$$f_{\mathbf{y}}(y) = \int_{\underline{x}} f_{\mathbf{y}|\mathbf{x}}(y|x) \cdot f_{\mathbf{x}}(x)\, dx \tag{C.19}$$

where $f_{\mathbf{y}|\mathbf{x}}(y|x)$ denotes the conditioned PDF of the event $\mathbf{y} = y$ under the assumption that $\mathbf{x} = x$ is observed. With the definition of the conditioned PDF and Equation C.19 it follows Bayes Theorem for continuous random variables

$$\begin{aligned} f_{\mathbf{x}|\mathbf{y}}(x|y) &= \frac{f_{\mathbf{x},\mathbf{y}}(x,y)}{f_{\mathbf{y}}(y)} \\ &= \frac{f_{\mathbf{y}|\mathbf{x}}(y|x) \cdot f_{\mathbf{x}}(x)}{f_{\mathbf{y}}(y)} \\ &= \frac{f_{\mathbf{y}|\mathbf{x}}(y|x) \cdot f_{\mathbf{x}}(x)}{\int_{\underline{x}} f_{\mathbf{y}|\mathbf{x}}(y|\widetilde{x}) \cdot f_{\mathbf{x}}(\widetilde{x}) d\widetilde{x}} \end{aligned} . \tag{C.20}$$

In order to derive the Bayes Theorem for mixed random variables, i.e. were $\mathbf{x}$ is a discrete random variable with the sample space $\underline{x} = \{x_0, x_1, \ldots, x_{N-1}\}$ and $\mathbf{y}$ is a continuous random variable, we use the delta function to represent the PDF the conditioned PDF of $\mathbf{x}$:

$$\begin{aligned} f_{\mathbf{x}}(x) &= \textstyle\sum_{j=0}^{N-1} P_{\mathbf{x}}(x_j) \cdot \delta(x - x_j) \\ f_{\mathbf{x}|\mathbf{y}}(x|y) &= \textstyle\sum_{j=0}^{N-1} P_{\mathbf{x}|y}(x_j|y) \cdot \delta(x - x_j) \end{aligned} .$$

Substituting these equations in Equation C.20 yields

$$\sum_{j=0}^{N-1} P_{\mathbf{x}|y}(x_j|y) \cdot \delta(x - x_j) = \frac{f_{\mathbf{y}|\mathbf{x}}(y|x) \cdot \sum_{j=0}^{N-1} P_{\mathbf{x}}(x_j) \cdot \delta(x - x_j)}{f_{\mathbf{y}}(y)}$$

and with $x = x_i$

$$\int_{x_i-\varepsilon}^{x_i+\varepsilon} P_{\mathbf{x}|y}(x_i|y) \cdot \delta(x - x_i)\, dx = \int_{x_i-\varepsilon}^{x_i+\varepsilon} \frac{f_{\mathbf{y}|\mathbf{x}}(y|x_i) \cdot P_{\mathbf{x}}(x_i) \cdot \delta(x - x_i)}{f_{\mathbf{y}}(y)} dx$$

we finally derive Bayes Theorem for mixed random variables

$$P_{\mathbf{x}|y}(x_i|y) = \frac{f_{\mathbf{y}|\mathbf{x}}(y|x_i) \cdot P_{\mathbf{x}}(x_i)}{f_{\mathbf{y}}(y)}. \tag{C.21}$$

C.8.2 Derivation of Equation 5.59

Equation 5.54 states

$$\begin{aligned} P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right) &= \frac{\sum_{b\in\underline{b^{(i)}}} e^{-\frac{E_S}{N_o}\cdot|r-g_b(b)|^2}}{\sum_{b\in\underline{b}} e^{-\frac{E_S}{N_o}\cdot|r-g_b(b)|^2}} \\ &= \frac{\sum_{s\in g_b\left(\underline{b^{(i)}}\right)} e^{-\frac{E_S}{N_o}\cdot|r-s|^2}}{\sum_{s\in\underline{s}} e^{-\frac{E_S}{N_o}\cdot|r-s|^2}}. \end{aligned} \tag{C.22}$$

With the prerequisites made on page 154, the set $g_b\left(\underline{b^{(i)}}\right)$ is either the union of $2^{\frac{n}{2}-1}$ different columns or rows. From now on we assume the first case - the proof for a union of rows is likewise. Hence,

$$g_b\left(\underline{b^{(i)}}\right) = \bigcup_{j\in\underline{p^{(i)}}} \underline{s_C^{(j)}},$$

where $\underline{p^{(i)}}$ and its complement depict the equal size dual partition of the set of all row numbers associated with the i-th bit. Therefore, the sum in the numerator of Equation C.22 can be ordered into a sum over the columns $\underline{s_C^{(j)}}$ associated with this bit and the sum in the denominator can be ordered into a sum over all columns:

$$\begin{aligned} P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right) &= \frac{\sum_{s\in\bigcup_{j\in\underline{p^{(i)}}}\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot|r-s|^2}}{\sum_{s\in\bigcup_{j=0}^{2^{\frac{n}{2}-1}}\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot|r-s|^2}} \\ &= \frac{\sum_{j\in\underline{p^{(i)}}}\left[\sum_{s\in\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot|r-s|^2}\right]}{\sum_{j=0}^{2^{\frac{n}{2}-1}}\left[\sum_{s\in\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot|r-s|^2}\right]} \\ &= \frac{\sum_{j\in\underline{p^{(i)}}}\left[\sum_{s\in\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot\left[(Re\{r\}-Re\{s\})^2+(Im\{r\}-Im\{s\})^2\right]}\right]}{\sum_{j=0}^{2^{\frac{n}{2}-1}}\left[\sum_{s\in\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot\left[(Re\{r\}-Re\{s\})^2+(Im\{r\}-Im\{s\})^2\right]}\right]}. \end{aligned}$$

For all symbols s in a specific column $\underline{s_C^{(j)}}$ the term $\left(Re\left\{r\right\}-Re\left\{s\right\}\right)^2$ is constant. If $x_j^2\left(r\right)$ denotes the squared distance from the j-th column to the receive symbol r, then

$$P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right) = \frac{\sum_{j\in\underline{p^{(i)}}}\left[e^{-\frac{E_S}{N_0}\cdot x_j^2(r)}\cdot\sum_{s\in\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot(Im\{r\}-Im\{s\})^2}\right]}{\sum_{j=0}^{2^{\frac{n}{2}-1}}\left[e^{-\frac{E_S}{N_0}\cdot x_j^2(r)}\cdot\sum_{s\in\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot(Im\{r\}-Im\{s\})^2}\right]}.$$

The sum $\sum_{s\in\underline{s_C^{(j)}}} e^{-\frac{E_S}{N_o}\cdot(Im\{r\}-Im\{s\})^2}$ represents a summation along the column j, which is a

constant for all columns. With K_y denoting this number,

$$\begin{aligned} P_{\mathbf{b}_i|\mathbf{r}}\left(1\,|r\right) &= \frac{\sum_{j\in\underline{p^{(i)}}}\left[e^{-\frac{E_S}{N_0}\cdot x_j^2(r)}\cdot K_y\right]}{\sum_{j=0}^{2^{\frac{n}{2}}-1}\left[e^{-\frac{E_S}{N_0}\cdot x_j^2(r)}\cdot K_y\right]} \\ &= \frac{K_y\cdot\sum_{j\in\underline{p^{(i)}}}e^{-\frac{E_S}{N_0}\cdot x_j^2(r)}}{K_y\cdot\sum_{j=0}^{2^{\frac{n}{2}}-1}e^{-\frac{E_S}{N_0}\cdot x_j^2(r)}} \\ &= \frac{\sum_{j\in\underline{p^{(i)}}}e^{-\frac{E_S}{N_0}\cdot x_j^2(r)}}{\sum_{j=0}^{2^{\frac{n}{2}}-1}e^{-\frac{E_S}{N_0}\cdot x_j^2(r)}}. \end{aligned}$$

Bibliography

[Abr65] M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions, National Bureau of Standards, Applied Math. Series #55, Dover Publications, 1965, sec. 6.5.

[Bos98] M. Bossert, Kanalcodierung, B.G. Teubner, Stuttgart, D, 1998.

[BCJR74] L. R. Bahl, J. Cocke, F. Jelinek, J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," IEEE Transactions on Information Theory, Vol. IT-20, pp. 284-287, March 1974.

[BGT93] G. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correction Coding and Decoding: Turbo Codes," IEEE Proc. ICC'93, pp. 1064-1070, Geneva, Switzerland, May 1993.

[Bnl87] G. Benelli, "An ARQ Scheme with Memory and Integrated Modulation," IEEE Transactions on Communications, Vol. COM-35, No.7, pp. 689-697, July 1987.

[Bnl92] G. Benelli, "A New Method for the Integration of Modulation and Channel Coding in an ARQ Protocol," IEEE Transaction on Communications, Vol. 40, No. 10, pp. 1594-1606, October 1992.

[Ben64] R. J. Benice, "An Analysis of Retransmission Systems," IEEE Transactions on Communication Systems, pp. 135-145, December 1964.

[Bre59] D. G. Brennan, "Linear Diversity Combining Techniques," Proceedings of the IRE, Vol. 47, pp. 1075-1102, June 1959.

[Bro95] I. N. Bronstein, K. A. Semendjajew, G. Musiol, H. Mühlig, Taschenbuch der Mathematik, Verlag Harri Deutsch, Frankfurt am Main, Germany, 1995.

[Co207] Failli M. (Chairman), Digital Land Mobile Radio Communications, COST 207 Final Report, Commision of the European Communities, September 1988.

[Cha56] S. S. L. Chang, "Information Feedback Systems," IEEE Transactions on Information Theory, pp. 29-40, September 1956.

[Cha61] S. S. L. Chang, "Improvement of two-way communication by means of feedback," 1961 IRE International Convention Record, Pt. 4, pp. 88-104, 1961.

[Chs85] D. Chase, "Code Combining A maximum-likelihood decoding approach for combining an arbitrary number of packets," IEEE Transactions on Communications, Vol. COM-22, pp. 385-393, May 1985.

[Dor86] B. G. Dorsch, "Adaptive Forward Error Correction for Channels with Feedback," System Research Center Report, University of Maryland, USA, August 1986.

[DVB97] ETSI, Digital Video Broadcasting (DVB); Multipoint Video Distribution System (MVDS) at 10 GHz and above, Technical Specification EN 300 748 V1.1.2, European Telecommunications Standards Institute, ETSI Secretariat, 06921 Sophia Antipolis Cedex, France, August 1997.

[For70] G. D. Forney, Jr., "Convolutional Codes I: Algebraic Structure," IEEE Transactions on Information Theory, Vol. IT-25, pp. 720-738, November 1970.

[Fri95] B. Friedrichs, Kanalcodierung: Grundlagen und Anwendungen in modernen Kommunikationssystemen, Springer, Berlin, D, 1995.

[Hag88] J. Hagenauer, "Rate-Compatible Punctured Convolutional Codes (RCPC Codes) and Their Applications," IEEE Transactions on Communications, Vol. 36, No. 4, pp. 389-400, April 1988.

[Har94] B. A. Harvey and S. B. Wicker, "Packet Combining Systems Based on the Viterbi Decoder," IEEE Transactions on Communications, Vol. 42, No. 2/3/4, February/March/April 1994.

[Hor63] M. Horstein, "Sequential Transmission Using Noiseless Feedback," IEEE Transactions on Information Theory, pp. 136-143, July 1963.

[Hoe90] P. Höher, Kohärenter Empfang trelliscodierter PSK-Signale auf frequenzselektiven Mobilfunkkanälen - Entzerrung, Decodierung und Kanalparameterschätzung, VDI Fortschrittberichte, Reihe 10, Nr. 147, VDI Verlag, Düsseldorf, 1990.

[Kal90] S. Kallel, "Analysis of a Type II Hybrid ARQ Scheme with Code Combining," IEEE Transactions on Communications, Vol. 38, No. 8, pp. 1133-1137, August 1990.

[Kal95] S. Kallel, "Complementary Puncture Convolutional Codes (CPC) and Their Applications," IEEE Transactions on Communications, Vol. 43, No. 6, pp. 2005-2009, June 1995.

[Kre89] U. H.-G. Kreßel, Informationstheoretische Beurteilung digitaler Übertragungsverfahren mit Hilfe des Fehlerexponenten, VDI Verlag, Düsseldorf, 1989.

[Kuh63] T. G. Kuhn, "Retransmission Error Control,", IEEE Transcations on Communication Systems, Vol. CS-11, pp. 186-201, June 1963

[Lin83] S. Lin, D. J. Costello, Jr., Error Control Coding: Fundamentals and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[LiYu82] S. Lin and P. S. Yu, "A Hybrid ARQ Scheme with Parity Retransmission for Error Control of Satellite Channels," IEEE Transactions on Communications, Vol. COM-30, No. 7, pp. 1701-1719, July 1982.

[Man74] D. M. Mandelbaum, "An Adaptive-Feedback Coding Scheme Using Incremental Redundancy", IEEE Transactions on Information Theory, Vol. IT-20, No. 3, pp. 388-389, May 1974.

[May98] A. Mayer, Improved Retransmission Strategies for ARQ Systems using Multidimensional Trellis Coded Modulation," Master Thesis, Technische Universität Darmstadt, 1998.

[McEl87] R. J. McEliece, Finite Fields for Computer Scientists and Engineers, Kluwer Academic Publishers, Dordrecht, NL, 1987.

[Met79] J. J. Metzner, "Improvements in Block-Retransmission Schemes", IEEE Transactions on Communications, Vol. COM-27, No. 2, pp. 524-532, February 1979.

[Mor78] J. M. Morris, "On another Go-Back-N ARQ Technique for High Error Rate Conditions," IEEE Transactions on Communications, Vol. COM-26, No.1, January 1978.

[Mor79] J. M. Morris, "Optimal Blocklengths for ARQ Error Control Schemes," IEEE Transactions on Communications, Vol. COM-27, No.2, pp. 488-493, February 1979.

[Nes63] M. Nesenbergs, "Comparison of the 3-out-of-7 ARQ with Bose-Chaudhuri-Hocquenghem Coding-Systems,", IEEE Transactions on Communication Systems, Vol. CS-11, pp. 202-212, June 1963.

[Opp96] A. V. Oppenheim, A. S. Willsky, S. H. Nawab, Signals & Systems, Prentice Hall, Upper Saddle River, NJ, 1996.

[Pal95] R. Palazzo, Jr., "A Network Flow Approach to Convolutional Codes," IEEE Transactions on Communications, Vol. COM-43, pp. 1429-1440, February / March / April 1995.

[PrGr58] R. Price and P. E. Green, Jr., "A Communication Technique for Multipath Channels," Proc. IRE, Vol.46, pp. 555-570, March 1958.

[Pro95] J. G. Proakis, Digital Communications, McGraw-Hill, New York, NY, 1995.

[Rap96] T. S. Rappaport, Wireless Communications: Principles & Practice, Prentice Hall, Upper Saddle River, NJ, 1996.

[Roc70] E. Y. Rocher and R. L. Pickholtz, "An Analysis of the Effectiveness of Hybrid Transmission Schemes," ICM Journal of Research and Development, pp. 426-433, July 1970.

[Sas75] A. R. K. Sastry, "Improving Automatic Repeat-Request (ARQ) Performance on Satellite Channels Under High Error Rate Conditions," IEEE Transactions on Communications, Vol. COM-23, No. 4, April 1975.

[Sas76] A. R. K. Sastry and L. N. Kanal, "Hybrid Error Control Using retransmission and Generalized Burst-Trapping Codes," IEEE Transactions on Communications, Vol. COM-24, pp. 385-393, April 1976.

[Schm98] M. P. Schmitt, A Novel ARQ Scheme Employing Adaptive Multiple/Multidimensional Trellis Coded Modulation, Master Thesis, State University of New York at Buffalo, NY, USA, 1998.

[Schm99] M. P. Schmitt, "Hybrid ARQ Scheme Employing TCM and Packet Combining ," IEE Electronic Letters, Vol. 34, pp. 1725-1726, September 1998.

[Schw63] L. S. Schwartz, "Feedback for Error Control and Two-Way Communication," IEEE Transcations on Communication Systems, Vol. CS-11, pp. 49-56, March 1963.

[Sha48] C. E. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal, Vol. 27, pp. 379 - 423, 623-656, July, October, 1948.

[Sin77] P. S. Sindhu, "Retransmission Error Control with Memory," IEEE Transactions on Communications, Vol. COM-25, No. 5, pp. 473-479, May 1977

[Sta94] H. Stark, J. W. Woods, Probability, Random Processes, and Estimation Theory for Engineers, Prentice Hall, Englewood Cliffs, NJ, 1994.

[Tan96] A. S. Tanenbaum, Computer Networks, 3rd Edition, Prentice Hall, Upper Saddle River, NJ, 1996.

[Wic95] S. B. Wicker, Error Control Systems for Digital Communication and Storage, Prentice Hall, Upper Saddle River, NJ, 1995.

[UMTS98] ETSI, Universal Mobile Telecommunications System (UMTS); Selection Procedure for the Choice of Radio Transmission Technologies of the UMTS; Technical Report 101 112, Sophia Antipolis Cedex, France, 1998 .

[Ung82] G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals," IEEE Transactions on Information Theory, Vol. IT-28, No. 1, pp. 55 - 67, January 1982.

[Yac93] M. D. Yacoub, Foundation of Radio Engineering, CRC Press, 1993.

Index